Meta-Analysis of Adolescent Drug Prevention Programs: Results of the 1993 Meta-Analysis

Nancy S. Tobler

INTRODUCTION

Policy relevant conclusions emerge when meta-analytic techniques are used to achieve consensus out of the inconsistencies found in individual research studies. Extensive search procedures located 120 school-based drug prevention programs that evaluated success on self-reported drug use measures. Hypothesis tests were made of an a priori classification scheme for the type of program. Six types of programs were identified based on content and delivery and were collapsed into noninteractive and interactive programs.

Because programs varied from 20 to 6,000 students, both ordinary least squares regressions (unweighted effect size) and weighted least squares regressions (weighted effect size) were conducted. Six covariates were entered into the regressions: sample size, targeted drug, type of control group, special populations, type of leader, and attrition. The relationship between program content, delivery, and the size of the programs was examined.

Interactive programs were significantly superior to the noninteractive programs in their ability to impact drug use behaviors and were equally successful for tobacco, alcohol, marijuana, and other illicit drugs. The effectiveness of the interactive programs was not only replicated, but increased with a subset of 56 high-quality experimental programs.

A meta-analysis of 143 adolescent drug prevention programs was completed by the author in 1986 and was reported elsewhere (Tobler 1986, 1992*b*. Tobler (1992*b*) is a reanalysis of 91 programs (a subset of the original 143 programs) that measured change solely on drug use outcome measures.[1] This chapter is based on a second data set which includes 120 adolescent drug prevention programs. One publication (Tobler 1993) emphasizes substantive material and gives a thorough description of the types of programs. Two publications report the methodology, inferential statistics, and the major findings (Tobler 1992*a*, 1994). In Tobler (1994) the data were reanalyzed to verify the major findings in Tobler (1992*a*), using a reduced set of relevant

variables as covariates so that the number of parameters is more in line with the number of cases.[2] This reanalysis is summarized briefly in this chapter.

For purposes of brevity, the two different meta-analyses are called 1986 and 1993. A number of differences should be noted. First, in 1986 the type[3] of prevention program was determined on a single dimension: the content or subject matter of the program. In 1993, this was expanded to include two dimensions: the content or subject matter and how the program content is delivered.[4] Second, the 1993 sample of drug prevention programs was limited to school-based prevention programs, whereas the 1986 sample included both school and community-based programs. Third, as adolescent drug use peaked in 1978, the 1993 meta-analysis examined only 1978 to 1990 data (versus 1972 to 1984 in the 1986 meta-analysis). This choice was made to reflect the downward societal trends in drug use (Johnston et al. 1986, 1989). Fourth, the 120 programs in the final 1993 sample all used drug use measures, versus 91 programs in the 1986 meta-analysis. The final set of programs in 1993 included 81 programs identified after the 1986 meta-analysis and 39 from the previous 1986 meta-analysis.

Finally, in 1993 the newest meta-analytical methodology was used to avoid the potential problem of arriving at incorrect conclusions due to inappropriate statistical procedures (Hedges and Olkin 1985; Hunter and Schmidt 1990; Rosenthal 1986). The 1993 results include both the unweighted effect size (UNES) and the weighted effect sizes (WES). Glass and colleagues (1981) defined UNES as the standardized mean difference between the treatment and the control group:

$$ES = (X_e - X_c)/SD_c, \quad (1)$$

where ES = effect size, X_e and X_c are the means for the experimental and control group, respectively, and SD_c is the standard deviation (SD) of the control group. In drug prevention research, parametric statistics are reported[5] which are computed using the pooled SD. To keep effect sizes comparable, it is more appropriate to use statistics which use the pooled SD, such as Cohen's d or its equivalent Hedges' g. The WES is then computed by weighting each effect size by the inverse of the variance, an estimate of the sample size. Hedges' (1986, p. 739) formula for the weighting factor of an individual study is:

$$W_i = [2(n_{ei} + n_{ci})n_{ei}\, n_{ci}] / [2(n_{ei}+n_{ci})^2 + n_{ei}\, n_{ci}\, d_i^2], \quad (2)$$

where W_i = weighting factor of the study, d_i = unweighted effect size, n_{ei} = number in the experimental group, and n_{ci} = number in the control group. Use of WESs is based on the fact that larger samples produce more stable results.

SELECTION CRITERIA

Selection Criteria for 1993

Criteria for inclusion in the 1993 meta-analysis were: (a) school-based drug prevention programs available to all members of the student body (may have included but did not target high-risk youth[6]); (b) reporting of drug use outcome measures; (c) use of a control or comparison group (comparison groups must have both pretest and posttest); (d) grades 6 to 12 (5th grade if incorporated into a middle school and/or longitudinal research was conducted); (e) goals of primary prevention, secondary prevention, and/or early intervention (does not target identified abusive/ compulsive or addicted drug users in treatment[7]); (f) participation of all ethnic groups that comprise the school's population; (g) location in United States and/or Canada; and (h) reported or published after 1977.

Additional Criteria for a Subset of Higher Quality Experimental Studies

The selection of a special subset of programs was made for two reasons. The first reason was to replicate the results with a set of solely experimental studies obtained from the mixed set of experimental and quasi-experimental studies. Many researchers feel that results of programs evaluated with quasi-experimental research designs yield overestimates of program effects; therefore, the analysis of a set of experimental studies will empirically examine this question. Second, as factors other than random or nonrandom assignment can impact evaluation results, a program was chosen that: (a) had a delivery intensity of not less than 4 hours (i.e., 1 week of classes); (b) administered a posttest not less than 3 months after pretest; (c) was not a placebo program even if the placebo program was compared to a control group (i.e., a program with one or more essential components deliberately excluded such as refusal skills); (d) was not compared to another

treatment program, (e) had followed individuals in longitudinal research (i.e., no cross-sectional research); and (f) had a measure of control for preexisting differences even if these differences were reported as nonsignificant (i.e., effect sizes could be computed from a change score, covariance adjusted means, or the individual's level of drug use at pretest).

META-ANALYTIC METHODOLOGY

Coding Procedures

A 50-page codebook was compiled that included over 250 variables related to: (a) treatment components (see table 1); (b) participant characteristics (e.g., grade, sex, ethnicity, socioeconomic class); (c) program characteristics (e.g., year, source of publication, goal, targeted drug, funding, location, number involved, number tested, research center); (d) implementation factors (e.g., intensity, duration, boosters, leaders, hours and type of leader training); (e) research methodology (e.g., sampling, assignment, unit of assignment, type of control group, research design, threats to internal validity); (f) test instrumentation (e.g., reliability, test-retest, internal consistency, reactivity of measure); and (g) data analysis (e.g., unit of data analysis, method of effect size calculation). In coding studies, the main focus was on gaining as much information as possible about the programs. If information was missing in the primary report or ambiguities needed clarification, researchers were contacted or additional literature searches were initiated. The principal investigator and two research associates independently coded all the content items. Ambiguous coding interpretations became the topic of discussion in the 2-hour weekly meetings and misinterpretations or errors were corrected.

A second "Manual for Effect Size Calculations" was developed for converting each of the summary statistics encountered (see Tobler 1992*a*, appendix 3). The principal investigator and two doctoral research associates, working independently from those coding content items, conferred about the choice of outcome measures and statistical procedures to use in calculating the effect size. Calculations were aided by a special computer software program (Tobler 1992*a*) and were spotchecked by the principal investigator.

Analysis—A Program

A program is the unit of analysis. In meta-analysis, studies are most often the unit of analysis with one effect size being reported per study (Bangert-Drowns 1986). But in drug prevention program research, some studies (i.e., research projects) compared the efficacy of more than one type of program. As the type of program is the variable of interest, using the study as the unit of analysis would not allow comparisons about the type of program. For example, "a cognitive program, a decisionmaking program and a values-clarification program" were compared in a single experimental study reported by Goodstadt and Sheppard (1983, p. 362). The three different types of alcohol education programs were administered to independent groups of adolescents, thereby contributing three effect sizes, one for each program type.

It was also necessary to insure that only one effect size was contributed to the overall analyses of a single program and a single group of adolescents. Numerous articles or reports were written about a single program. Each of the articles related different information about the same program such as results for different testing periods (i.e., pretest information, immediate posttest, and followups). Often details about the program content, instrumentation, and implementation were included in separate publications. To insure independence of a sample of students, all authors were cross-checked against all other authors in the database to identify duplicate reports on the same group. Sets of articles or reports were then sequenced by pretest, posttest, and followup results and given one program number.

Independence of Outcome Measure

Each outcome measure category estimated the effect of the program based on a different concept. If two or more effect sizes on the same outcome measure were reported for a program, they were averaged and recorded as one effect size. Using this procedure, a student was represented only once in a specific outcome measure category. As results were not averaged across outcome categories, a student could not be represented more than once in the overall analysis for that outcome measure.

Every outcome measure reported at baseline was traced through all testing periods. Frequently, a large number of these measures were not reported in the final results. It was assumed that failure to report on all of the initial measures indicated nonsignificant findings and an effect of zero was assigned, a conservative method.

Independence for Type of Drug

Effect sizes were kept independently for five categories of drugs: cigarettes, alcohol, marijuana, hard drugs (cocaine, heroin, stimulants, inhalants, and tranquilizers), and "all drugs." The "all drugs" category accommodated programs with various combinations of drugs not reported separately. If more than one effect was reported for a category, the mean was reported as a single effect for that category. Each category was kept independently to facilitate later analyses by type of drug. For the main analyses—one effect per program—the results were averaged across types of the drugs. Behavioral intentions were not included as a drug use measure.

Independence for Subpopulations

If results were broken out separately by sex, grade, and/or level of drug use (nonuser, experimental user, user), individual effect sizes were calculated. For example, if three types of outcome measures were reported for boys and girls for three levels of drug use, 18 effects were computed (3 outcomes x 2 sexes x 3 levels). "Because...different students are involved in each of these comparisons, the effect sizes derived from the comparisons are independent" (Giaconia and Hedges 1982, p. 585). To obtain one program effect for the final analysis, the effect size for each subpopulation was averaged. For example, in a program having a positive effect for the boys and a negative effect for girls, the mean effect for the program is zero and does not accurately portray the program's results. Bangert-Drown's (1986) study effect method (one effect per program) does not take into account differential results across subpopulations. Because the WES was used, the weighting factors for the individual subpopulations were also combined into a single weighting factor for the program. But, in this case, the sums of the individual subpopulation weights were computed to be used at the aggregate level (see Tobler 1994).

Pooling Effect Sizes Over Test Intervals for a Single Program

Effect sizes were computed for each subpopulation for all testing periods reported. The exact number of months from pretest to posttest and/or followup was coded. A categorical variable was created: (a) 1 to 12 months, (b) 13 to 24 months, (c) 25 to 36 months, and (d) greater than 37 months. If more than one test was given in an interval, the average was reported. This occurred frequently in the first time interval as many programs gave a posttest and followup test within 12 months. None of the time

intervals included all of the programs, so it was necessary to consider pooling effects across test intervals. However, analyses were first conducted to determine if effects decreased or increased with time. Three statistical procedures were used. First, a repeated-measures multivariate analysis of variance (MANOVA) was found to be nonsignificant for programs (N = 4) with results in all four time periods. A second repeated-measures MANOVA for programs (N = 12) in the first and the fourth time intervals was also found nonsignificant. Further inspection showed that equal numbers of programs reported increases in effect size over time as those reporting decreases in effect size over time. Third, scatterplots of 118 programs[8] compared each time period with each other. The scatterplots also supported the pooling of effects sizes (for greater detail, see Tobler 1992*a*).

A second aggregation produced a final single effect for a program by averaging the effects for the time intervals reported (Tobler 1994). This method maintains the statistical independence for each program.

Choice of Covariate Adjusted Means

Effect sizes are usually computed on the final unadjusted posttest results (Glass et al. 1981; Smith et al. 1980). Unadjusted means can only be used when random assignment resulted in truly equivalent treatment and control groups. Undoing the covariate adjusted scores to obtain the unadjusted means, as proposed by Smith and colleagues (1980) and Glass and colleagues (1981), would remove all the control built into the data analysis to correct for the problem of preexisting differences. In fact, the best-designed programs that initially blocked on preexisting drug use would be penalized the most. As the purpose of meta-analysis is to show program effects, not preexisting differences, the program effect sizes were computed from the covariate adjusted means reported by the researcher. Also, including quasi-experimental (nonrandom assignment) studies necessitates working with change scores; an assumption of no preexisting differences between groups at pretest cannot be made. Additionally, the unit of random assignment for experimental programs was intact social units, either classrooms (27 percent) or schools (53 percent), rather than individuals (27 percent). Only 43 percent of those studies randomly assigning intact units had more than six experimental and six control units, which leaves preexisting differences a major problem. As a final consideration, test-retest reliabilities are needed to compute unadjusted posttest scores whether analysis of covariance summary statistics or pretest/posttest means and SDs are available for effect size

computations. Test-retest values were not reported in 81 percent of the studies in Tobler (1992*a*). Convention rules for estimating test-retest reliabilities were developed by Smith and colleagues (1980); however, these are gross estimates, either underestimating or overestimating the actual effect size.

Windsorizing

Based on a precedent set by Lipsey (1992) in a meta-analysis of juvenile delinquency treatment, a decision was made to windsorize the weighting factor. This was accomplished by limiting the weighting factor for the larger programs to a maximum and increasing the weighting factor of the smaller programs. This decision was necessary as the sample of students in a program varied from 20 to about 6,000. The weighting factor is the inverse of the variance, which is approximately four times smaller than the number of participants in the program. Twenty-one programs had weighting factors under 25 (100 tested students or less), while six programs had weighting factors near or above 1,000 (i.e., 4,000 tested students). Without windsorizing, the largest programs would be given 40 times the weight of the smaller programs, allowing one large study to completely overshadow the results of the smaller programs. To reduce the 40:1 ratio to a more reasonable 8:1 ratio, the weighting factors under 30 were windsorized up to 30 and the larger programs over 250 were limited to 250. The number present at each test was used to determine the weighting factor.

Other Decisions

When frequencies, proportions, or percentages were the only data reported, probit transformations (Cohen and Cohen 1983) were used to compute the effect size. The use of probit transformations with change scores is discussed elsewhere (Tobler 1985). Where parametric statistics were reported, the effect sizes were calculated using documented formulas (Tobler 1992*a*, appendix 3). When reports stated that results were significant, a 0.05 level of significance was assumed and the corresponding t levels computed. If only a statement of nonsignificance was reported, a p value of 0.50 was assigned (i.e., an effect size of zero). This is a conservative method for estimation of effect sizes. Had researchers given the actual p value, even though not significant, it would lead to an effect size greater than zero.

INDEPENDENT VARIABLE—TYPE OF PROGRAM

Program Content

Program content was coded for 30 items and collapsed into 7 major domains: knowledge, affective, drug refusal skills, generic skills, safety skills, extracurricular activities, and others (see table 1). The content items were coded either yes or no; therefore, the coding scheme did not reflect the relative time spent on a particular content area. This necessitated a subjective decision about the amount of emphasis placed on a particular content item before categorization.

Group Process

The methods and techniques used to deliver the program content have been given little emphasis in the review literature. The terms "interactive" and "noninteractive" were chosen to emphasize what actually happened in the classroom. The type of group process or delivery method was incorporated into the definition of the type of program in the 1993 scheme. "When we observe how a group is handling its communication, i.e., who talks how much or who talks to whom, we are focusing on group process" (Edwards 1972, pp. 182-183). As drug prevention programs are carried out with either the whole class or in smaller groups, a group classification system based upon Toseland and Rivas's topology (1984, pp. 20-22) was specifically revised to describe the classroom processes operating in school-based drug prevention programs (table 2). Four types of groups (A to D) were identified and were ranked on a continuum, beginning with group A (table 2, left column) which had little or no adolescent interaction (i.e., didactic presentations). Each group included progressively greater degrees of interaction among the group members, with group D being the most interactive.

Determination of the Type of Program

Once the decisions about the content and type of group were made, the two dimensions were combined to determine the type of program. Using items listed in table 3, a decision was made about which combination of the five major content domains (knowledge, affective, refusal skills, generic skills, and safety skills[9]) best portrayed the program content. The second choice was the type of group (right column, table 3). The overall context of the entire program was taken into consideration before making a final determination about the type of program. For example, for Drug

TABLE 1. *Major content in adolescent drug prevention programs.*

KNOWLEDGE
Knowledge of drug effects
Knowledge of media and social influences
Knowledge of actual drug use by peers (normative education)
AFFECTIVE
Self-esteem and feelings
Personal insight and self-awareness
Attitudes, beliefs, and values
REFUSAL SKILLS
Drug-related refusal skills
Public commitment activities
Cognitive behavioral skills
Support systems/networking with nondrug-using adolescents
GENERIC SKILLS
Communication skills
Assertiveness skills
Decisions/problemsolving skills
Coping skills
Social/dating skills
Goal-setting
Identifying alternatives
SAFETY SKILLS
Skills to protect self in a drug-related situation
Skills to protect other peers in a drug-related situation
Drinking/driving safety
EXTRACURRICULAR ACTIVITIES
Paid job activities or training
Organized sports
Organized cultural activities
Nondrug leisure time activities
Volunteer work in the community
OTHER
Peer counseling/facilitating/helping
Homework exercises
Rewards, token economy, and reinforcement
Parent involvement
Communitywide coordination and involvement

SOURCE: Reprinted with permission: Tobler, N. Updated meta-analysis of adolescent drug prevention programs. In: Montoya, C.; Ringwalt, C.; Ryan, B; and Zimmerman, eds. *Evaluating School-Linked Prevention Strategies: Alcohol, Tobacco and Other Drugs*. San Diego: USCD Extension, University of California, 1993. pp. 71-86.

TABLE 2. *Four group types.*

	Group A	Group B	Group C	Group D
Aim	To educate: knowledge gain	To educate: intrapersonal competence; self-awareness, self-esteem building, feelings, values, "affective education"	To develop: interpersonal skills; relationships with others; increase feeling of acceptance through positive peer interactions	To develop: intrapersonal and/or interpersonal growth.
Purpose	Learning through didactic presentation	Some didactic presentations; group discussions; individually oriented experiential activities	To increase communications and social skills; improve interpersonal relationships through structured exercises, role plays; and interpersonal experiential activities	To identify member's potentials; self-awareness, insight, and interpersonal development through discussion and dynamic group process
Leadership	Leader as teacher	Leader as teacher; provider of structure for group	Leader as facilitator of the group's activities; provide structure	Leader as facilitator and role model; group members take responsibility for group's direction
Focus	Individual knowledge	Individual growth	Focus on group as a medium for interaction; involvement of all individuals	Either member or group focus; individual growth through the group experience
Structure	Highly structured; passive participation	Structured; passive and some active participation	Structured; active participation	Limited structure; open ended, active participation

TABLE 2. *Four group types (continued).*

	Group A	Group B	Group C	Group D
Bond	None	Common interest in learning; skills development; bond limited	A common activity, enterprise or situation; bond between members	Common goals among members; contract to use the group to grow; bond between members
Composition	Typical school class	Similarity of educational or skill level	Can be diverse or homogeneous	Can be quite diverse. Based on members' ability to work towards growth and development
Communication patterns	Leader to student	Leader to member, didactic; sometimes member to member during discussions; self-disclosure low	Member to member; often represented in activity and nonverbal behavior; self-disclosure low to moderate	Highly interactive; members often take responsibility for communication in the groups; self-disclosure moderate to high

SOURCE:Reprinted with permission: Tobler, N. Updated meta-analysis of adolescent drug prevention programs. In: Montoya, C.; Ringwalt, C.; Ryan, B.; and Zimmerman, R., eds. *Evaluating School-Linked Prevention Strategies: Alcohol, Tobacco and Other Drugs.* San Diego: USCD Extension, University of California, 1993. pp. 71-86.

Awareness Resistance Education programs (Project DARE), a determination was made that relatively more emphasis was placed on intrapersonal content in the affective and generic domains than on the interpersonal skills such as drug refusal (Ringwalt et al. 1990). Project DARE's content would be coded knowledge, affective, some refusal skills, and some generic skills, and was classified as knowledge-plus-affective because the emphasis on refusal and interpersonal skills was limited. The second choice, the group process, also placed Project DARE under knowledge-plus-affective as it was most typically delivered in a noninteractive group B setting. Twenty-six distinct types of programs were identified, consolidated into the six major

subcategories, and further collapsed into two overarching categories: noninteractive programs and
interactive programs.

DEPENDENT VARIABLE

The drug use outcome measures were paper-and-pencil self-reports given confidentially in a classroom setting and were often accompanied by physical tests (i.e., saliva). The reliability of confidential self-reports of cigarette use has been documented (Murray et al. 1987; O'Malley et al. 1983; Pechacek et al. 1984) as have measures of other illicit and licit drug use (Oetting and Beauvias 1990; Single et al. 1975).

DATA ANALYSIS

Ordinary and Weighted Least Squares Regression

Ordinary least squares (OLS) regression analyses were used for the unweighted effect size. For the weighted effect size, weighted least squares (WLS) regression analyses were conducted as detailed in Hedges and Olkin (1985). This procedure weights each program effect size by the sample size of that program. The significance testing is conducted at the program level (SPSS 1990).

To account for the differences in the effectiveness of a type of program, other variables related to program success must be considered. For example, recent smoking programs have been highly successful and the possibility exists that their success is the result of targeting cigarettes and not the type of program used. Multiple regression procedures make available methods for computing the unconfounded effect for the type of program by partialing out the effect of all the covariates (i.e., holding constant the effect of the covariates). A discussion of each covariate is included in the following sections.

Dummy Coding for Categorical Variables

In the present analyses, the dependent variable (effect size) and one covariate (sample size) are continuous variables. The remaining six predictor variables are categorical. The independent variable, type of program, is categorical, as are the five covariates: type of control group, experimental design, special populations, targeted drug(s), and

TABLE 3. *Type of program by content and process.*

Content	Process
NONINTERACTIVE: KNOWLEDGE ONLY	
Knowledge	Group A
Knowledge	Film/theater
Knowledge + Attitudes	Group A
Drinking + Driving	Group A
Drinking + Driving	Scare tactics
AFFECTIVE ONLY	
Affective	Group B *ECM
Affective	Group B
KNOWLEDGE PLUS AFFECTIVE	
Knowledge + Affective	Group B
Knowledge + Affective + Attitudes + Values	Group B
Knowledge + Affective + Decisions	Group B
Knowledge + Affective + Generic	Group B
Knowledge + Affective + Refusal + Generic	Group B
Knowledge + Affective + Generic + Community	Group B
Drinking + Driving	Group B
INTERACTIVE: SOCIAL INFLUENCES	
Knowledge + Refusal	Group C
Knowledge + Refusal + Community**	Group C
Drinking + Driving	Group C
COMPREHENSIVE LIFE SKILLS	
Knowledge + Refusal + Generic	Group C
Knowledge + Refusal + Generic + Community**	Group C
Drinking + Driving	Group C
OTHERS	
Knowledge + Norm-changing	Group C
Knowledge + Affective	Group C
Knowledge + Affective	Group D
Knowledge + Affective + Generic	Group C
Knowledge + Affective + Refusal + Generic	Group D

KEY: * = Effective classroom management for teachers; ** = Total community effort supporting the school-based program.

SOURCE:Reprinted with permission: Tobler, N. Updated meta-analysis of adolescent drug prevention programs. In: Montoya, C.; Ringwalt, C.; Ryan, B.; and Zimmerman, R., eds. *Evaluating School-Linked Prevention Strategies: Alcohol, Tobacco and Other Drugs.* San Diego: UCSD Extension, University of California, 1993. pp. 71-86.

leaders. The type of program or independent variable is comprised of two clusters of programs: noninteractive and interactive. Therefore, it was dummy coded, 1 or 0, to identify group membership. Three other covariates were comprised of binary clusters: type of control group, experimental design, and special populations. Two covariate variables were comprised of a cluster of more than two dummy variables. For example, leaders consisted of a cluster of four different types of leaders: teachers, same age or older age peer leaders, mental health professionals, and all others. Teachers were designated as the reference group and were coded 0, 0, 0. The peer leaders were coded 1, 0, 0; mental health professionals were coded 0, 1, 0; and all others were coded 0, 0, 1. In dummy coding, the degrees of freedom for a variable are K-1; therefore, a binary variable uses one degree of freedom. Three degrees of freedom are used for the leaders variable, which is composed of a cluster of four dummy variables.

Regression Equation

To examine the effects due to the primary independent variable (type of program) without the confounding effects of the covariates, it was necessary to remove the proportion of variance attributed by each covariate. Each of the covariate clusters was entered into the regression equation before the primary independent variable. The sequence of entry for the covariates was arbitrary as no order was hypothesized. The effects of the six confounding covariates were removed before computing the covariate adjusted means for two types of programs.

To keep the number of parameters in line with the number of cases, interactions were not included. Partial confirmation for this is given by the fact that the two-way analyses of variance (ANOVAs) for each covariate with the primary independent variable had no significant second-order interaction effects. Finally, the OLS residuals were examined for outliers. Six outliers were identified and removed, leaving a sample of 114 programs.

Of interest is the extent that a covariate accounts for program success. It is important to answer questions such as, "Which is more highly associated with program success, the type of program or the drug targeted by the program?" The increment to R^2, which is the proportion of variance accounted for by a covariate, can be used to determine the relative impor-tance of a variable for predicting program efficacy. No attempt was made to independently analyze any of the levels within the categorical covariates. For the primary independent variable, the magnitude of the change in R^2 can be determined when this variable is

entered into an equation that already contains the covariates (i.e., partialing out the effect of all the covariates).

Hypothesized Covariates Omitted

The variables identified as potent predictors of program success were chosen based on previous research (Tobler 1986) and a review of the literature. Sex, initial level of drug use, booster sessions, implementation factors, and the research center were all eliminated as covariates because only a limited number of programs reported results broken out for this information (frequencies are reported in Tobler 1992*a*).

Two additional hypothesized variables, grade and program intensity, were eliminated based on the analyses reported in Tobler (1992*a*). Each variable was nonsignificant in all 16 regression analyses and contributed R^2 increments of less than 2 percent.

Six Covariates Included

Sample Size. The effect sizes for the programs with large sample sizes were found to be smaller (Tobler 1992*a*); therefore, the weighting factor, which is an approximate estimate of the sample size, was entered as a continuous variable.

Type of Control Group. Treatments compared to a no-treatment control group were found to have higher effect sizes than those compared to a standard health curriculum/another treatment (Tobler 1986, 1992*a*). The reference category was treatments compared to a health class control.

Experimental Design. A categorical variable was made for studies that had acceptable attrition (with or without differential dropout) and unacceptable attrition (with or without differential dropout). The reference category was acceptable attrition. This binary variable was derived from the empirical findings reported (Tobler 1992*a*). A decision tree was used which involved three choices: assignment, attrition, and differential dropout.

The results showed that no differences in effect sizes were observed for random (0.17) versus nonrandom (0.16) assignment. Whether differential dropout occurred from treatment or control was missing in 61.7 percent of the reports (Tobler 1992*a*); these studies were grouped with those reporting differential dropout (a conservative method). As a result of the complex empirical results for experimental design, it was decided that the experimental design was best represented by the binary variable of acceptable

and unacceptable attrition. The retention rates for school-based drug prevention studies were compiled as part of a meta-analysis of 85 longitudinally followed cohorts (Hansen et al. 1990). This data provided normative attrition rates for drug prevention research. Attrition was coded as acceptable if it was on the mean or above (12 months from pretest) and unacceptable if below the mean.

Special Populations. The literature reports that most research has been conducted primarily in schools with > 50 percent white populations. In Tobler (1992*a*), schools with > 50 percent minority or problem students were found significantly more successful than those with > 50 percent white populations in a number of regressions for the 114 programs. The reference category was schools with > 50 percent white populations.

Targeted Drug. Three categories existed for this dummy variable: smoking programs, alcohol programs, and substance abuse and/or generic drug prevention programs. The generic drug programs have outcome measures for cigarettes, alcohol, marijuana, and all other drugs. Therefore, the effect size must be seen as an average of the results for all drugs tested, whereas smoking and alcohol programs tested a single drug. It was not possible to examine the results for a single drug in the generic programs and still use the study effect method (one effect per program). The reference group was smoking programs.

Leaders. Four categories of leaders were entered for this block: teachers, peer leaders, mental health specialists,[10] and others (e.g., research staff, health educators, and various outside professionals). The reference group was teachers.

RESULTS

As the aim of this chapter is to connect the descriptive statistics and the qualitative information to the inferential analyses, this section briefly discusses the inferential statistics.[11] The mean effect sizes are presented for only the unadjusted means as the focus is descriptive. It should be noted that the difference between the unadjusted mean effect sizes and the covariate adjusted mean effect sizes was very small (approximately 0.01). The results are reported for both UNES and WES. The WES is meta-analytically sound as the effect size has been weighted for the sample size. The UNES is reported even though it is considered meta-analytically unsound, because it provides a way to examine the results for the smaller programs without their being overpowered by the larger programs. Two problems

specific to drug prevention program research makes this necessary: the wide range in sample sizes (20 to 4,000 tested youth), and the limited number of evaluated programs, which precludes separate meta-analyses for the smaller programs and larger programs.

Nature of Programs Located

The strength of a meta-analysis depends on the comprehensiveness of the sample of programs. Extensive search procedures located 595 studies of adolescent prevention programs. The 120 programs that passed the selection criteria came from 90 research studies: single programs were reported in 69 studies; more than one type of program was reported in 21 studies.[12]

Five hundred and five studies (84.9 percent) did not pass the selection criteria. Eighty-three studies (14.1 percent) that were evaluated lacked a control or comparison group; 113 studies (19.2 percent) were descriptive in nature with subjective conclusions. Another group of 30 (5.1 percent) studies was excluded after passing the original selection criteria. These 30 studies initially appeared to have an experimental or quasi-experimental research design, but on further reading it was found that they were not implemented according to the original plan and/or they lacked the necessary statistical data to calculate an effect size.

Other studies were excluded as follows: 132 were not school based and/or did not target high-risk youth[13]; 45 evaluated grades under fifth or college students; 30 did not have drug use outcome measures; 21 were published earlier than 1978; 13 were published before 1978 and also did not have drug use measures; 17 were not implemented in the United States or Canada; 2 evaluated only the teachers; and finally, time did not allow the inclusion of 19 eligible studies.

The 143 drug prevention programs in the 1986 meta-analysis were included in the tally above. Only 39 programs passed the new selection criteria. Because comparisons will be made between the 1986 and 1993 meta-analyses, the reasons for excluding 104 of the 143 programs are also listed: 13 did not meet the more stringent research design qualifications,
13 lacked drug use measures, 29 were published before 1978, 24 were published before 1978 and also lacked drug use measures, 22 targeted high-risk youth and/or were community based, 2 were eliminated due to insufficient time, and 1 report was a duplicate of the same group (i.e., second posttest).

Effect Size by Type of Program

The 70 interactive school-based adolescent drug prevention programs were effective in changing adolescent drug use behaviors (UNES = 0.247, 0.95 confidence interval (CI) = 0.18 to

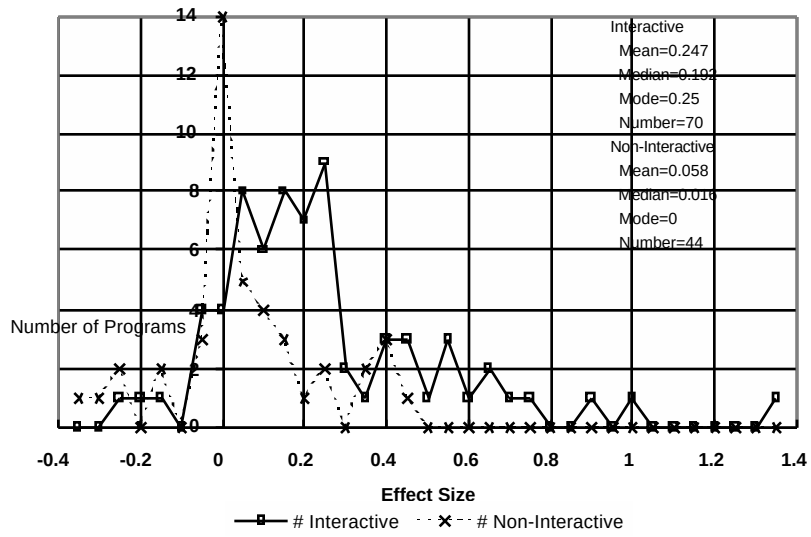
31; WES = 0.164, 0.95 CI = 0.14 to 18) while the 44 noninteractive programs were essentially ineffective (UNES = 0.058, 0.95 CI = 0.00 to 0.11, WES = 0.075, 0.95 CI = 0.05 to 0.10). A further analysis of the subset of 56 high-quality experimental programs included in the larger set of 114 programs[14] showed even higher effect sizes for the interactive programs (UNES = 0.317, 0.95 CI = 0.22 to 0.41; WES = 0.214, 0.95 CI = 0.19 to 0.24; N = 38) and still lower effect sizes for the noninteractive programs (UNES = 0.017, 0.95 CI = -0.07 to 0.11; WES = 0.043, 0.95 CI = 0.00 to 0.09; N = 18). In all four regressions, the interactive programs were significantly better than the noninteractive programs (P = 0.002 for the unweighted OLS for 114 programs; P = 0.001 for the unweighted OLS for 56 programs; P = 0.009 for the weighted WLS for 114 programs; and P = 0.015 for the weighted WLS for 56 programs).

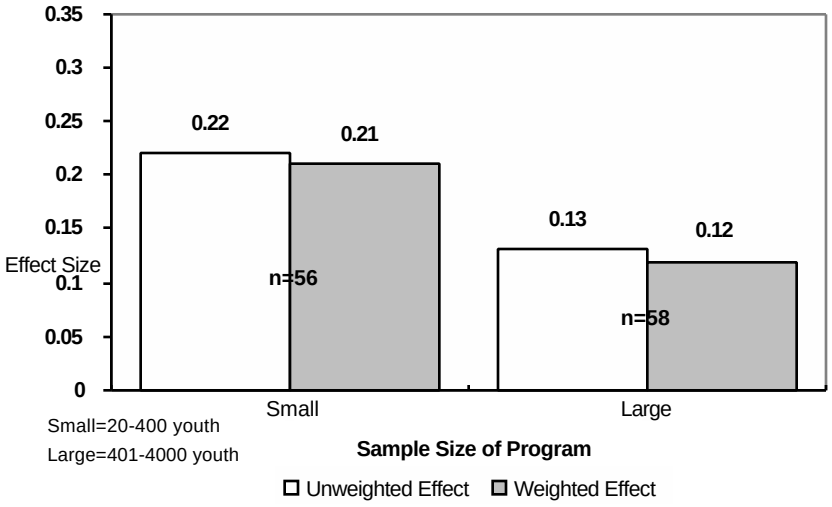
Effect Size Distribution by Noninteractive and Interactive Programs

Figure 1 gives the frequency distribution of the UNES for the two types of programs. The distributions for the interactive and noninteractive programs are shown separately and are strikingly different. The 44 noninteractive programs have a mean of 0.058, a median of 0.016, a mode of zero, a range of -0.35 to 0.45, and 0.95 CIs of 0.00 to 0.11. This stands in contrast to the 70 interactive programs which have a mean of 0.247, a median of 0.192, a mode of 0.25, a range of -0.24 to 1.34 and 0.95 CIs of 0.18 to 0.31. When the two separate distributions are compared to the combined distribution for all 114 programs, it can be seen that the noninteractive programs were responsible for the mode, while the interactive programs contribute more to the positive skew.

Effect Size by Sample Size

There is a limitation, however. All the large scale implementations (i.e., 400 to 4,000 tested youth), whether interactive or noninteractive programs, exhibited a leveling of effectiveness (see figure 2). The smaller programs (i.e., 20 to 400 tested youth) had an UNES of 0.22 (0.95 CI = 0.14 to 0.31) and a WES of 0.21 (0.95 CI = 0.18 to 0.25). The large scale programs had an UNES of 0.13 (0.95 CI = 0.08 to 0.17) and a WES of 0.12 (0.95 CI = 0.10 to 0.14). This is a ds of 0.09.


Interactive
Mean=0.247
Median=0.193
Mode=0.25
Number=70
Non-Interactive
Mean=0.058
Median=0.016
Mode=0
Number=44
Number of Programs
14
12
10
8
6
4
2
0
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
1.2
1.4
Effect Size
Interactive
Non-Interactive


0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Effect Size
0.22
0.21
0.13
0.12
n=56
n=58
Small
Large
Small=20-400 youth
Large=401-4000 youth
Sample Size of Program
Unweighted Effect
Weighted Effect

Effect Size by Posttest Time Interval

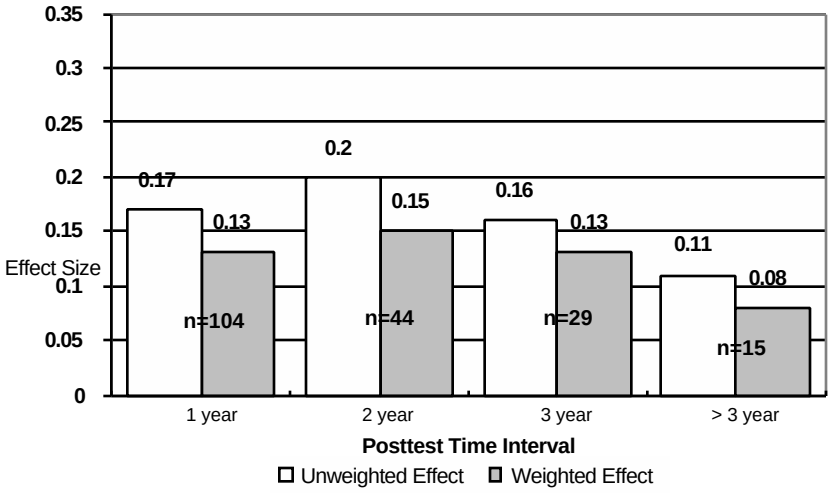
The results at each of the four test intervals (1 year, 2 years, 3 years, and > 3 years) are given in figure 3. The magnitude of the effect sizes were maintained over the first 3 years and showed a slight decrease for the fourth interval (greater than 3 years). Most probably, the minor variations across time were due to the different sets of programs included in each interval. No single time interval included all the programs. Ninety percent of the programs reported test results within the first year; a sharp drop was observed for the second year, with only 34 percent of programs represented; 25 percent reported third year results; and only 15 percent took posttests at a period greater than 3 years. The total number does not equal 114 because many programs administered multiple posttests.

To alleviate concerns about decay of program effectiveness over time, the four OLS and WLS regressions were rerun using the first posttest results as the dependent measure regardless of the length of time from pretest. To control for effectiveness decay over time, the length of time from pretest to the first posttest was entered as an additional continuous covariate along with the original six covariates.

The results of OLS and WLS regression based on the first posttest were almost identical to those based on the average effect size across time intervals. Using the first posttest results, the interactive programs were significantly better than the noninteractive programs: $P = 0.003$ (first posttest) versus $P = 0.002$ (across time) for the unweighted OLS for 114 programs; $P = 0.005$ (first posttest) versus 0.009 (across time) for the weighted WLS for 114 programs; $P = 0.002$ (first posttest) versus $P = 0.001$ (across time) for the unweighted OLS for 56 programs; and $P = 0.015$ (first posttest) versus 0.015 (across time) for the weighted WLS for 56 programs. The effect size across time and the effect size for the first posttest showed identical patterns for the interactive and noninteractive programs. The only noteworthy observation was that the noninteractive programs were underrepresented for the second year, third year, and the fourth interval (greater than 3 years).

Effectiveness of the Six Major Subgroups

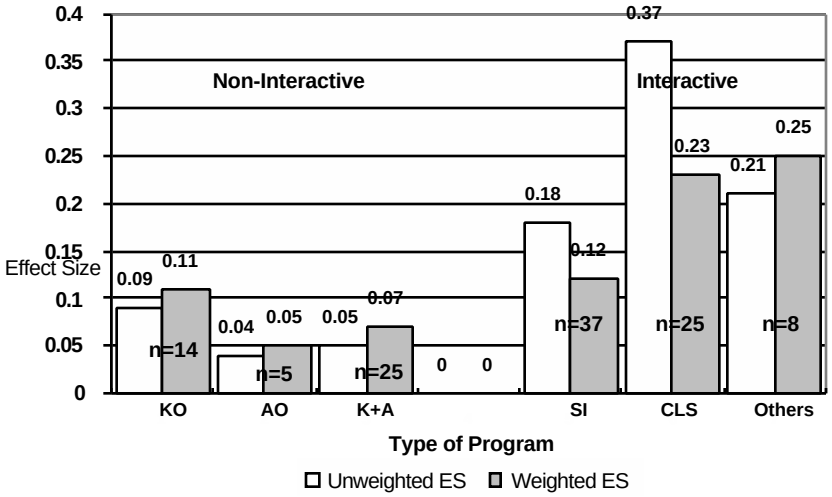
Set of 114 Programs. The noninteractive programs included three subgroups: knowledge only (KO), affective only (AO), and knowledge-plus-affective (K+A) programs (see figure 4). The three interactive programs subgroups were social influences (SI), comprehensive life skills



(CLS), and others (see right side of figure 4). In the set of 114 programs, the highest effect sizes were obtained by the CLS programs. The others programs had the second highest effect sizes, but also the largest confidence interval. The third most effective subgroup was the SI programs, which also had the tightest confidence intervals.

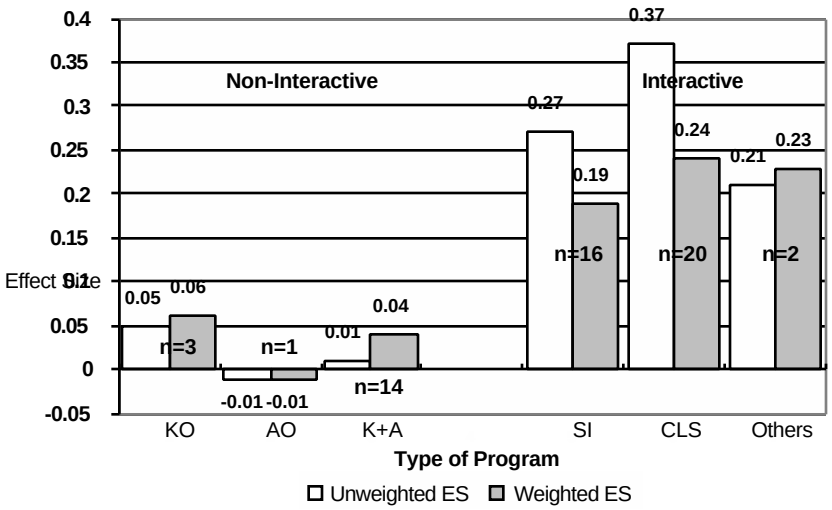
All three subgroups of interactive programs had higher effect sizes than the three subgroups of noninteractive programs.

Subset of 56 Programs. Differences were observed when comparing the 114 programs (figure 4) with the subset of 56 programs (figure 5). For the interactive programs, the effect sizes for the SI programs were substantially higher in the subset of 56 programs than in the 114 programs, while the CLS and others programs remained about the same. The three noninteractive programs showed lower effect sizes for the well-controlled programs (subset of 56).



One Type of Program Eliminated: Drinking/Driving

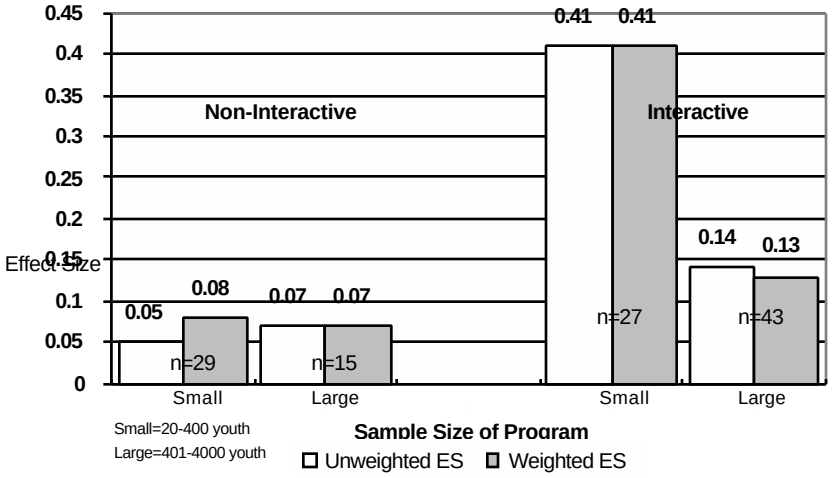
It was hypothesized that drinking/driving programs should constitute a major type of program based solely on the program content (Goodstadt, personal communication, November 10, 1989). These programs emphasize an individual's responsibility in alcohol-related accidents or deaths, personal safety relative to driving with someone who has been drinking, and responsibility for providing safe transportation for friends who have been drinking (e.g., designated drivers). Only 10 drinking/driving programs were located. Within this subgroup, there were five different types of programs: KO, KO with fear tactics, K+A, SI , and CLS programs. Surprisingly, the effect sizes ranged from -0.18 to 0.30, showing extreme heterogeneity; this indicates that other factors were operating besides program content. When reclassified by the program delivery method (group A to D), the effect sizes matched those of programs having similar delivery methods (i.e., same types of groups). The limited number of drinking/driving programs precludes any definitive analyses, but the fact that the empirical results show that the type of group is more important than the content in determining the type of program validates the inclusion of the type of group process in the classification
scheme.



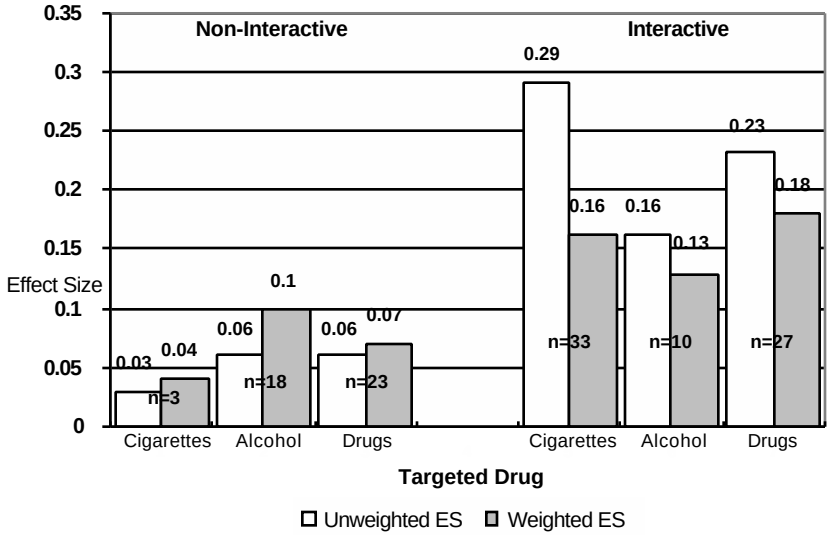
Six Covariates by Type of Program

The UNES and WES were presented for only the larger set of 114 programs in order to maintain enough programs in a category. Without knowing the distribution for each of the categories, the interpretation of a category mean with a limited number of programs becomes very tenuous. For example, if 9 programs show small but consistent positive effects while the 10th program is highly negative, the resulting mean may be zero or even slightly negative. Therefore, the results for the set of 56 programs are described only when they differ from the 114 programs.

Sample Size. The sample size, a continuous covariate, was found significant in all four regressions reported in Tobler (1994), as was the independent variable, type of program. However, the proportion of variance accounted for by sample size was lower than the independent variable, but much higher than any of the other five covariates. When broken into groups of small or large size programs within the noninteractive and interactive categories, the small interactive programs were extremely successful (figure 6). Both their UNES and WES were 0.41, whereas the small noninteractive programs achieved only an UNES of 0.05 and WES of 0.08. A large drop in effectiveness was observed when the small, highly successful interactive programs were implemented on a larger scale. Still, the large interactive programs were twice as successful as large noninteractive programs. (Notice that the UNES and WES were nearly equivalent in each of the four groups.) For the remaining covariates, the WES will be much lower, reflecting the loss of effectiveness when implemented on a large scale.



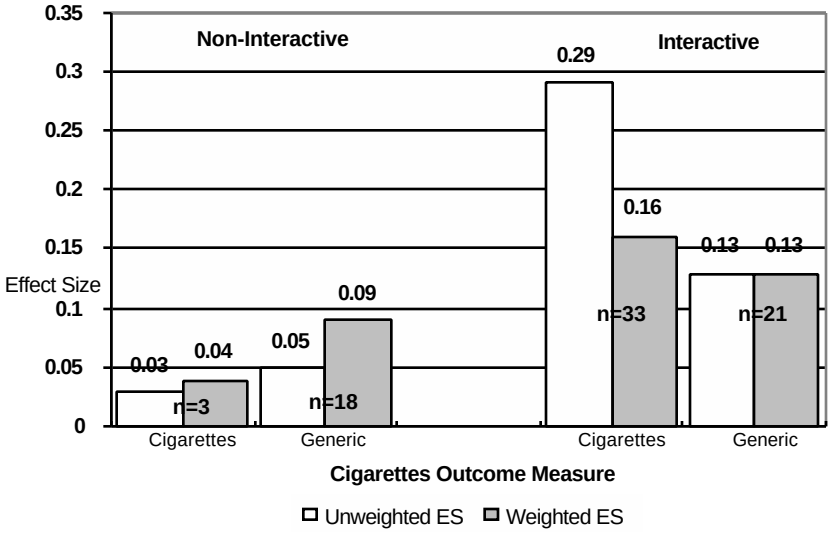
Targeted Drug. Figure 7 shows the results of smoking, alcohol, and generic drug prevention programs by type of program. This covariate was significant in both of the unweighted OLS regressions (56 and 114 programs), which indicates that the smaller smoking programs were more effective than the larger generic drug prevention programs. Relative to the size of the programs, the interactive smoking programs had an UNES almost twice as high as its WES. Relative to targeting a drug, the interactive programs had an UNES for the smoking programs that was larger than either the alcohol or the generic drug programs; all the WESs were about the same. In other words, the smaller programs were more effective in all cases, although the differentials between the UNES and WES were not as large for the alcohol programs or the generic drug programs. For the noninteractive programs by program size, a smaller differential was observed between the UNES and WES and it was in the opposite direction. The differences between the smoking, alcohol, and generic programs were minimal.

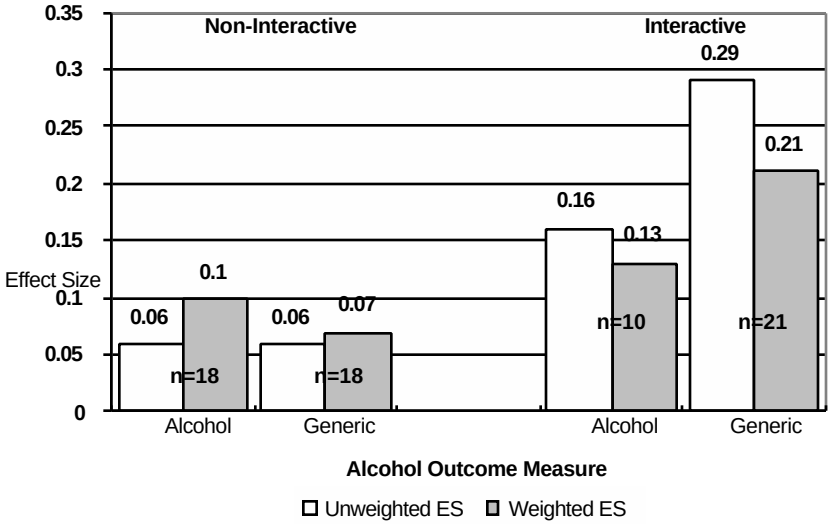


Generic Drug Programs. The interactive programs were nearly four times as effective as the noninteractive programs for UNES size and about three times as effective for WES. Note that generic drug programs tested for cigarettes, alcohol, marijuana, and all other illicit drugs; a composite score was necessary to maintain one effect size per program. Furthermore, no independent analysis was made of the levels of dummy coded variables within each categorical covariate (i.e., between smoking, alcohol, and generic programs). Therefore, further analyses were necessary to determine the effectiveness of the generic drug programs with cigarettes and with alcohol. To accomplish this, the cigarette score was extracted from the composite score (i.e., mean for cigarettes, alcohol, marijuana, and other drugs) reported by the generic programs and then compared to the results for the smoking programs (figure 8). A similar procedure was used for alcohol (figure 9).

Smoking Programs. Whether a smoking program or a drug prevention program, the interactive programs were significantly superior to the non-interactive programs ($Q_B = 7.95$, $Q_{cr} = 3.841$) in reducing cigarette use. Therefore, any further comparisons must be made within the categories of interactive or noninteractive programs. Within the 54 interactive programs, the 33 smoking programs were not significantly superior to the 21 generic drug prevention programs ($Q_B = 1.62$, $Q_{cr} = 3.841$). The small interactive smoking programs were extremely beneficial, as seen in the UNES magnitude (figure 15). However, the WES for the interactive smoking programs was not much better than the interactive generic programs, although both were higher than the noninteractive programs. The reverse was true for the generic noninteractive programs, which had a higher WES than UNES, but only three noninteractive programs targeted cigarettes. However, these measurements were performed on the set of 114 programs with

all their inherent problems; therefore, to validate these results a further analysis was made using the 56 high-quality experimental programs.[15]





Unfortunately, only 38 programs tested cigarettes in the subset of 56 programs. The interactive programs had a sufficient number of programs (i.e., 14 smoking and 15 generic) for comparison; the noninteractive programs had only 1 smoking and 8 generic programs. The interactive smoking programs were highly successful and significantly superior to the interactive generic programs ($Q_B = 33.38$, $Q_{cr} = 3.841$). The effect sizes for the smoking programs (UNES =

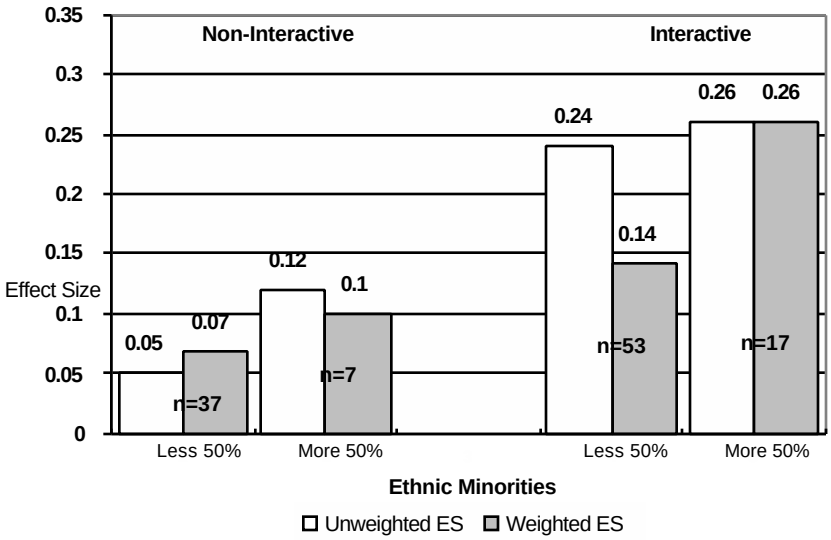
0.48, WES = 0.32) were much higher than those obtained for cigarettes in the generic programs (UNES = 0.11, WES = 0.12). The higher UNES versus WES for the smoking programs indicates that the smaller smoking programs were more successful than the larger smoking programs. Apparently the size of the program was not a factor for the generic drug programs, as the UNES and WES were almost equivalent.

Alcohol Programs. The interactive programs were significantly superior to the noninteractive programs ($Q_B = 23.42$, $Q_{cr} = 3.841$) for alcohol use based on the alcohol outcome measure for the set of 114 programs (figure 9). For alcohol programs, it should be noted that only 32 percent of the interactive programs targeted alcohol versus 50 percent of the noninteractive programs. Within the noninteractive programs there were no significant differences between programs that targeted alcohol and the generic approaches ($Q_B = 0.98$, $Q_{cr} = 3.841$); the interactive programs showed that generic programs were slightly superior to alcohol programs ($Q_B = 4.670$, $Q_{cr} = 3.841$).

Again, because of the potential sources of bias in the set of 114 programs, the subset of 56 high-quality experimental programs was examined. The interactive programs were superior to the noninteractive programs ($Q_B = 20.6$, $Q_{cr} = 3.841$). Only one of the four subcategories, interactive generic programs, was large enough for reliable conclusions; the noninteractive alcohol, the noninteractive generic, and the interactive alcohol programs had less than 10 cases, which makes these categories vulnerable to spurious findings. Surprisingly, the generic interactive programs achieved an effect size for alcohol (UNES = 0.29, WES = 0.21, N = 15) that was approximately twice as high as the results for cigarettes using generic interactive programs.

Special Populations. The F change for schools having greater than 50 percent ethnic minorities was statistically significant in only one regression: WLS regression for 114 programs (P = 0.009). The UNES and WES were identical (0.26) for the interactive programs in schools having greater than 50 percent minorities (figure 10); this similarity indicates that the larger programs produced results equivalent to the smaller programs. Within the interactive programs, the opposite was true for schools with less than 50 percent minorities. A much lower WES (difference of 0.10) than UNES indicates that when the interactive programs were implemented with a white student population on a large scale, they did not do as well. The noninteractive programs were also slightly more successful in schools with greater than 50 percent ethnic minorities.

Leaders. No statistically significant findings were obtained for the F change for leaders in any of the four regressions, and the increment to R^2 was below 2 percent. For the interactive programs, mental health specialists were the most effective and teachers were the least effective (figure 11). Same age/older age peer leaders and other professionals were slightly more successful than the teachers within the interactive programs. The pattern was the same within the noninteractive programs with the exception of mental health specialists, who were the least successful leaders (N = 4). A degree of confidence can be placed in these
findings



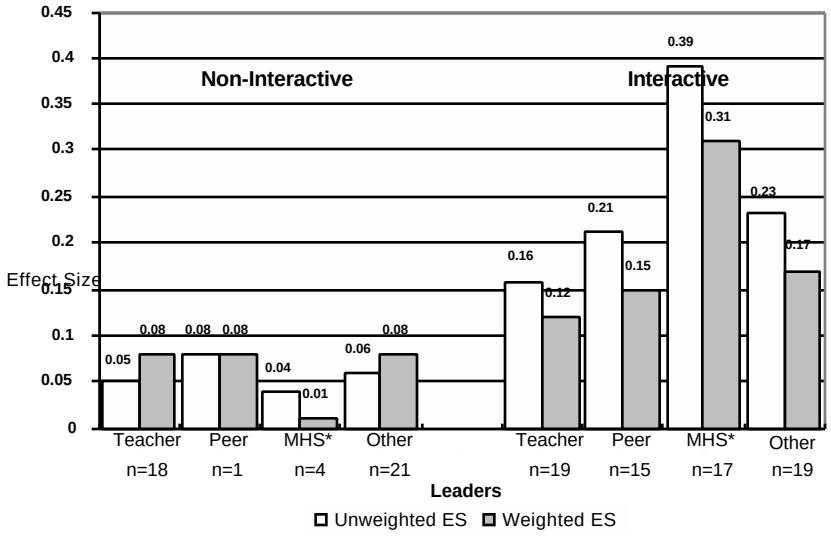
because the four types of leaders were almost equally represented within the interactive programs, providing an excellent opportunity for comparison.

Experimental Design. The F change for experimental design was not significant in any of the four regressions. Unacceptable attrition was only slightly related to smaller effect sizes within the interactive programs, while it accounted for very little difference in the noninteractive programs (figure 12).

Type of Control/Comparison Group

The F change for the type of control group did not reach significance in any of the regression equations, most probably because the mean differences (approximately 0.08) between the health class control and the no-treatment control were not large enough (figure 13). Examining the differences within the noninteractive and interactive programs shows equivalent mean differences between the two different types of control/comparison groups. Also, the two types of

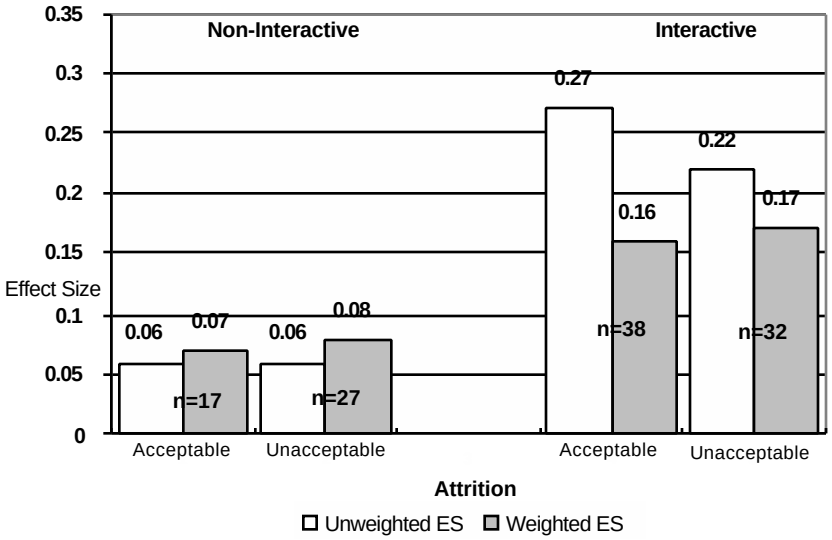
control/comparison groups occurred with the same frequency in both the noninteractive and interactive
programs.



Content and Process Components by Type of Program

The frequencies for the most prevalent content items and the type of group process are shown for the set of high-quality experimental studies (see table 4). Only three of the six major subgroups are included: K+A, SI, and CLS. These groups contained a sufficient number of programs to allow comparisons of their content and process components. The three remaining major subgroups were not included because of the limited numbers: KO (N = 3), AO (N = 1), and others (N = 2). Because the focus of this section is to identify specific content items associated with a major program type, four drinking/driving programs were also excluded as their content was much different.

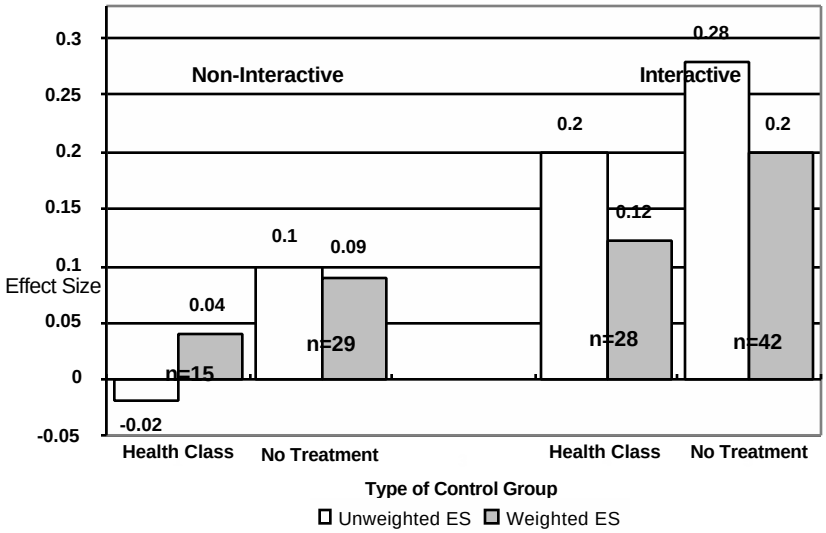
K+A Noninteractive Programs. Within the K+A category, two types of programs were identified and empirically confirmed by cluster analysis procedures. The two subcategories were called values and DARE type programs. Both the values and the DARE type programs include



knowledge about drug effects, a strong emphasis in the affective domain, and also delivery in a noninteractive setting (group B).

Values Programs. Media influences were included in some of the values programs, although none included information about normative expectations. In the affective domain the major emphasis was attitudes and values, followed by insight and self-awareness. None of the values programs included drug refusal skills, although they did emphasize generic decisions/problemsolving skills.

The values programs differed from the DARE type, SI, and CLS programs in the importance placed on insight and self-awareness for the purpose of changing attitudes and values. These components were nonexistent in the DARE type, SI, and CLS programs. Notably, the values programs did not include information about normative expectations or drug refusal skills; these were found in the SI and CLS programs and to some degree in the DARE type programs. Although decision/problemsolving skills were frequently used in the values programs, the focus was intrapersonal, not interpersonal



DARE Type Programs. All five DARE type programs included knowledge about drug effects, media influences, and normative expectations. A different emphasis was placed on the affective content than in the values programs. None of the DARE type programs incorporated self-awareness and insight, but all stressed self-esteem and feelings. Only 40 percent included attitudes and values. Drug refusal skills and decision/problemsolving skills were present in all the DARE type programs, although the amount of time spent on refusal techniques was limited. Another content domain emphasized was generic skills (e.g, assertiveness skills, coping skills).

DARE type program content was closely related to the CLS programs, particularly the inclusion of generic skills, but the DARE type programs did not emphasize drug refusal skills or other interpersonal skills as much as the CLS programs. The DARE type programs placed a greater emphasis on self-esteem than the CLS programs. The most important difference between the DARE type programs and CLS programs was the manner in which the content was delivered. The DARE type programs

TABLE 4. *Contents and process by type of program.*

	Noninteractive		Interactive	
	Values	DARE	SI	CLS
	N = 9	N = 5	N = 13	N = 19
KNOWLEDGE				
Knowledge of drug effects	100%	100%	100%	100%
Media & social influences	44%	100%	69%	63%
Normative expectations	0%	80%	77%	74%
AFFECTIVE				
Self-esteem, feelings	33%	100%	0%	53%
Insight, self-awareness	56%	0%	0%	0%
Attitudes, beliefs, & values	78%	40%	0%	5%
REFUSAL SKILLS				
Drug-related refusal skills	0%	100%	92%*	100%
Public commitment activities	11%	60%	46%	32%
GENERIC SKILLS				
Communication skills	0%	20%	8%	74%
Assertiveness skills	22%	60%	8%	63%
Decisions/problemsolving	78%	100%	8%	95%
Coping skills	11%	60%	8%	74%
Social/dating skills	0%	0%	31%	58%
Goal setting	33%	20%	0%	68%
SAFETY SKILLS				
Skills to protect peers	0%	0%	0%	0%
Drinking/driving safety	11%	0%	0%	0%
PROCESS				
Group A noninteractive	0%	0%	0%	0%
Group B noninteractive	100%	100%	0%	0%
Group C interactive	0%	0%	100%	100%
Group D interactive	0%	0%	0%	0%

KEY: * = Culturally sensitive to Native American population.

used a noninteractive group process versus the interactive group process used by the CLS programs.

SI and CLS Interactive Programs. All the SI and CLS programs include knowledge of drug effects, drug refusal skills, and an interactive setting (group C).

The SI programs were highly focused drug refusal skills programs with only two other content items: media influences and normative expectations. One program did not include refusal skills but had all the other SI components. This modified program was designed to be culturally sensitive for a Native American population. The programs that the Native American adolescents received were atypical and

presented classification problems, as was observed in the cluster analyses.

The CLS programs included media influences and normative expectations as frequently as the SI programs. However, as the name implies, CLS programs were more comprehensive. These programs included many generic skills that were not related solely to the use of drugs (i.e., decision/ problemsolving, communication skills, coping skills, goal setting, assertiveness skills, and social skills). In other words, the CLS programs subsumed all the components of the SI programs, added intrapersonal skills, and also included additional nondrug interpersonal skills.

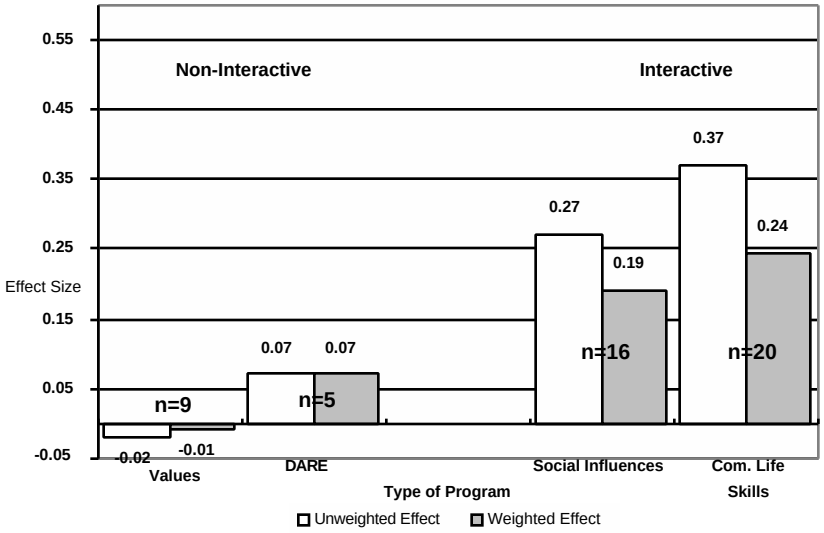
Effect Sizes for Values, DARE Type, SI, and CLS Programs

The magnitude of the UNES and WES increases from left to right in figure 14. The values programs were essentially zero. The DARE type programs had much lower effect sizes than either of the interactive programs. The SI programs were higher than the DARE type programs but lower than the CLS programs, which had the highest effect sizes.

Four of the five DARE type programs were Project DARE program evaluations that had been delivered to sixth graders. To alleviate concerns about being unable to detect program success with sixth graders (i.e., very low use rates), the four Project DARE evaluations were compared only to the sixth grade programs. As Project DARE shares content with both the noninteractive programs and the interactive programs, comparisons were made to both types of programs. The 16 fifth and sixth grade interactive programs effect sizes were much higher (UNES = 0.35, WES = 0.19) than the four Project DARE programs (UNES = 0.07, WES = 0.07) and were also higher than the other nine noninteractive programs (UNES = 0.05, WES = 0.08).

Comparisons Between Interactive Programs: SI Versus CLS

None of the regressions reported in Tobler (1992*a*) were statistically significant for the planned comparisons of SI programs and CLS programs. These results were confusing; the CLS programs effect sizes were consistently



higher than the SI effect sizes. Two possible confounds were discovered: more mental health specialists (the most successful leaders) conducted the CLS programs, and the SI programs were implemented more frequently on a large scale. First, to address the leaders issue, the OLS and WLS regressions were rerun with the regressions detailed in this chapter using only the 37 SI programs contrasted against the 25 CLS programs. Second, the regressions were repeated after eliminating all the programs that used mental health specialists as leaders.

The F change for the type of program (i.e., SI versus CLS) was nonsignificant in the four unweighted regressions (see table 5A) and the four weighted regressions (see table 5B). Additionally, the increment to R^2 for the SI versus CLS contrast accounted for less than 1 percent of the total R^2. Only one covariate, sample size, had a significant F change in all eight regressions. The increment to R^2 for sample size accounted for most of the total R^2. One other covariate—targeted drug—was significant, but only in the two OLS regressions.

TABLE 5A. *UNES and OLS regressions, SI versus CLS programs with and without MHS as leaders.*

	SI		CLS					
114 programs	N	X_{SI}	N	X_{CLS}	X_{CLS}-X_{SI}	R^2 change	F change	Sig. F
SI vs CLS	37	0.18	27	0.37	0.19	0.9%	0.881	0.352
Targeted drug						8.8%	4.458	0.016
Sample size						34.1%	31.026	0.000
56 PROGRAMS								
SI vs CLS	16	0.27	20	0.37	0.10	0.9%	0.451	0.508
Targeted drug						11.7%	3.110	0.060
Sample size						30.9%	15.176	0.000
114 PROGRAMS WITHOUT MHSs								
SI vs CLS	34	0.15	17	0.29	0.14	0.6%	0.426	0.518
Targeted drug						11.2%	4.196	0.021
Sample size						28.1%	19.173	0.000
56 PROGRAMS WITHOUT MHSs								
SI vs CLS	13	0.22	14	0.30	0.08	1.3%	0.392	0.540
Targeted drug						15.0%	2.600	0.999
Sample size						26.9%	9.186	0.006

TABLE 5B. *UNES and WLS regressions, SI versus CLS programs with and without MHS as leaders.*

	SI		CLS					
114 programs	N	X_{SI}	N	X_{CLS}	X_{CLS}-X_{SI}	R^2 change	F change	Sig. F
SI vs CLS	37	0.12	27	0.23	0.11	0.4%	0.355	0.554
Sample size						25.0%	20.020	0.000
56 PROGRAMS								
SI vs CLS	16	0.19	20	0.23	0.04	1.2%	0.490	0.491
Sample size						23.7%	10.562	0.003
114 PROGRAMS WITHOUT MHSs								
SI vs CLS	34	0.12	17	0.20	0.08	1.2%	0.078	0.781
Sample size						17.5%	10.417	0.002
56 PROGRAMS WITHOUT MHSs								
SI vs CLS	13	0.17	14	0.21	0.03	1.0%	0.252	0.622
Sample size						16.8%	5.034	0.034

Even though the effect sizes were lower after removing the mental health specialists, the large effect size differences between the SI and CLS programs remained (see tables 5A and 5B). Both the regressions and effect sizes indicate that the size of the program was the most important covariate when contrasting the SI and CLS programs.

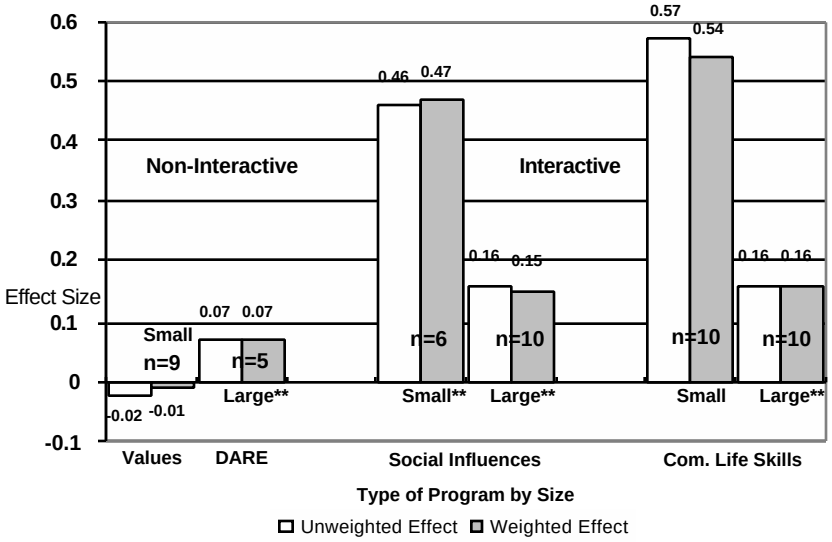
Effect Sizes by Sample Size for the K+A, SI, and CLS Programs

To further examine the relationship of program size to the type of program, the subset of 56 high-quality experimental studies was used as other areas of potential bias had also been eliminated. Only the K+A, SI, and CLS programs had enough cases to be further subdivided by size. Fortuitously, when the K+A programs were subdivided by size, all the small programs were in the subcategory called values and all the remaining large K+A programs were DARE type programs. Before examining the relationships between the six groups of programs (figure 15), it should be noted that the UNES and WES programs for each group were nearly identical. The large differences that existed between the UNES and WES for the other five covariates were not present when divided by the size of the program.

Both of the small interactive (SI and CLS) programs were extremely successful, while the small noninteractive (values) programs were totally ineffective. Also, the large mean differences between the SI programs and the CLS programs were substantially reduced; the small SI programs were nearly as effective as the small CLS programs. More important, the effect sizes for the large SI programs were equal to the large CLS programs. Unfortunately, both the large SI programs and the large CLS programs were only one-third as effective as their counterparts when implemented on a smaller scale. Despite this drop, the large interactive programs were still twice as effective as the large noninteractive DARE type programs.

Effectiveness by Drug Type for Noninteractive and Interactive Programs.

The differential results for the noninteractive and interactive programs by type of drug are presented in figure 16. The interactive programs were equally successful for cigarettes, alcohol, and marijuana; the UNESs ranged from 0.22 to 0.33 and the WESs ranged from 0.15 to 0.39. The effect size for marijuana was slightly higher than that for alcohol, with tobacco use having the lowest effect size of the three most frequently used drugs. Illicit drugs, excluding marijuana, had extremely high effect sizes but these results were based on only six programs.

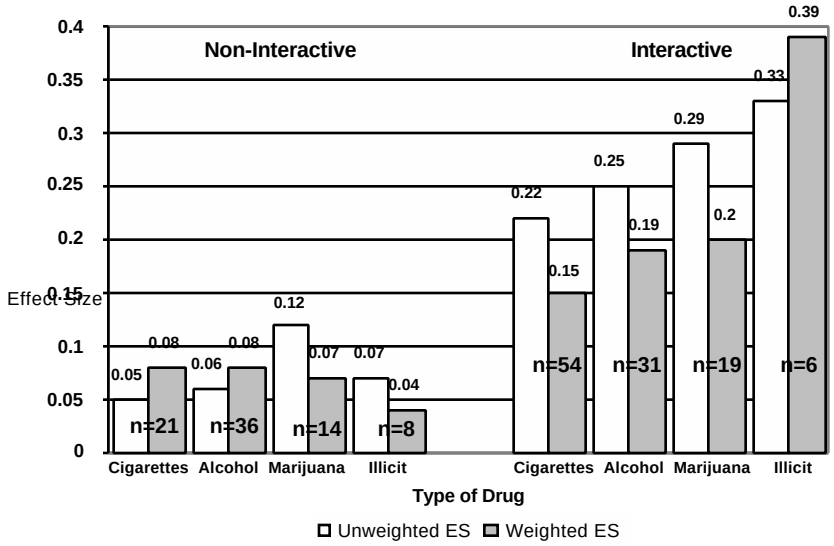


Further examination of the 56 high-quality programs showed effect sizes more closely related to other drugs: 0.19 UNES and 0.18 for the WES. The noniteractive programs were equally unsuccessful with all four types of drugs; the UNES ranged from 0.05 to 0.12, and the WES ranged from 0.04 to 0.08.

DISCUSSION

A Priori Organizational Scheme

Similar to primary research analyses, meta-analysis can be used to investigate relationships or to test specific hypotheses. The 1986 meta-analysis was exploratory; a wide net was cast to include a variety of programs and thereby identify relationships that were developed after a thorough coding of 143 programs. The relationships identified in 1986 laid the groundwork for the development of specific hypotheses, particularly concerning the type of program. Even though the 1993 analysis of type of program was based on two dimensions (content and process), it was similar to, and was a continuation of, the 1986 organizational scheme.



In the 1993 analysis, a program was coded and placed in one of the six subgroups in the predetermined classification scheme. Cluster analyses for 20 content and 4 process items verified the similarity of the programs within the 6 subgroups. The six subgroups were divided into the two major types of programs and were then tested with a priori planned comparisons (Tobler 1992*a*). When testing a specific hypothesis, the direction of inference is opposite of that found in an exploratory meta-analyses. "A hypothesis asserts which treatment is most effective: a review then examines empirical evidence to test the hypothesis" (Light and Pillemer 1984, p. 27).

Descriptive Analyses Confirm Two Types of Programs

The pessimistic reports of drug prevention program research have definitely resulted from the improper combining of two independent sets of programs. When analyzed collectively, the efficacy of drug prevention programming is questionable. Together, the mean effect size was 0.17 and the mode was zero. These results echo the pessimistic conclusions of the traditional literature reviews. However, when separated by the type of program (based on the a priori organizational scheme), two independent effect size distributions were observed. A second distribution for the effect size by sample size also indicated that two independent groups were combined. The noninteractive and interactive programs had similar funnel distributions showing a definite leveling of effect size, albeit at different magnitudes. A substantial difference was observed: the interactive programs

had a mean effect size of 0.25 and a mode of 0.25, whereas the noninteractive programs had a mean effect size of 0.06 and a mode of zero. Both distributions verify the need to analyze the two types of programs separately and clearly illustrate the danger in grouping all drug prevention programs into one category. Instead of arriving at the incorrect assumption that nothing works, it can be concluded that although not all drug prevention programs work, the interactive programs were effective.

Group Process

The largest effect size differences were found between the noninteractive and interactive programs. Substantively, the characteristic that specified the difference between the noninteractive and interactive programs was the method used to deliver the program's content (i.e., the group process). Irrespective of the program content, the noninteractive programs did not emphasize interactions between peers as did the more participatory interactive programs. In fact, the delivery method or group process, not the content, was fundamental in defining the two types of programs. The majority of programs had multiple content components (Hansen 1992; Tobler 1993), and these overlapped within and between the two major types of programs (tables 1 and 3). Because the group process was not only an integral part, but perhaps was central, in defining the noninter-active and interactive programs, a brief review follows (see table 2 for greater detail).

Noninteractive Group Process. The two noninteractive groups, A and B, used classroom dynamics familiar to all teachers. In the least interactive group (A), the leaders delivered a didactic presentation in a manner similar to a math, history, or health class. For the most part, these highly structured classes did not actively involve the students. Group B format was used by the majority of the noninteractive programs. Although a structured lecture format (i.e., passive) was used to present information, these groups also reported that students actively participated in teacher-led discussions. Experiential activities were incorporated, but these activities remained focused on the individual rather than on interactions with others in the group. For example, a values clarification exercise might involve adolescents independently listing their personal values, but the results of the exercise were generally shared only with the group leader in exchanges that excluded group peers.

Interactive Group Process. Interactive group process skills have been defined by the Office of Substance Abuse Prevention

(1989, p. xiv): "This teaching technique is used to stimulate active participation of all students in the classroom activity, be it discussion, brainstorming session, or practice of new behaviors." Optimally, in group C the interactions included everyone and were both participatory and between peers. Structured small group activities were used to introduce program content and promote the acquisition of skills. This highly structured format was developmentally appropriate for younger adolescents, who bond with their peers as they participate in activities together. The leader keeps the group on track by initiating appropriately timed structured activities. Ideally, all adolescents practiced their newly acquired skills and received corrective feedback in a supportive atmosphere, enabling them to use their new skills in a situation of higher stress (i.e., a real world, drug-related situation).

The second interactive group, group D, was the converse of the traditional classroom. Group D had the least structure and, therefore, was more appropriate for older adolescents. Only three programs reported using this type of group. Even these groups maintained a definite structure and were neither wide-open discussions nor therapy groups. Optimally, the leaders in both interactive groups encouraged everyone to participate, promoted positive and supportive interactions between the adolescents, and assumed an authoritative role only when it was necessary to correct a misconception.

Importance of Sample Size

The success of the interactive programs was not without a caveat: the loss in effectiveness demonstrated by the larger programs was disappointing. This post hoc finding was second in importance only to the a priori hypotheses about the type of program. Although the mean effect size differences between the small programs and large programs were not quite as large as those observed between the two types of programs, the size of the program was also statistically significant in all the regressions. The magnitude of these effect size differences mandates that comparisons be made between similarly sized programs. Ideally, two independent meta- analyses should be conducted; one for the smaller efficacy trials[16] and one for the larger scale effectiveness trials.[17] However, this approach was not possible in the present analysis because of the limited number of studies.

For the small programs in the set of 114 programs, the extraordinary superiority of the 27 interactive programs is

evident in figure 6. It is important to note that even when implemented on a small scale under ideal conditions, the noninteractive programs were ineffective. The difference between the small noninteractive and interactive programs was 0.36 for the UNES and 0.33 for the WES. When comparing the large programs, the differences between the noninteractive and interactive programs were much smaller, 0.07 for the UNES and 0.06 for the WES. Still, the large interactive programs were twice as effective the large noninteractive programs (figure 6).

Content

Focus of Noninteractive Programs.

The content of the noninteractive programs was directed towards individuals and their own internal perceptions, and therefore had a primarily intrapersonal focus. Despite variations within the noninteractive programs (KO, AO, and K+A subcategories), the program content maintained an intrapersonal focus. For example, the KO programs stressed the acquisition of factual knowledge about the physical and psychological consequences of drug use. The theoretical assumption was that given sufficient knowledge, the adolescent would develop negative drug attitudes that, in turn, would lead to healthy personal choices. The AO programs assumed that psychological factors place certain persons at risk of use and/or abuse. Various activities focused on building self-esteem and self-awareness, and promoting positive personal feelings with the aim of increasing the individual's intrapersonal competence and social functioning (no information about drugs was provided). The majority of the AO programs included in this meta-analysis trained teachers extensively in use of effective classroom management techniques (Moskowitz et al. 1984) for the purpose of altering the entire classroom milieu.

The K+A programs also had an intrapersonal focus, yet the two subcategories, value programs and DARE type programs, were based on very different theoretical assumptions. The values programs aimed to change the individual's personal attitudes and values about drug use. Therefore, the content included knowledge, decisionmaking skills, problemsolving skills, goal setting, values clarification, and so forth. These programs encouraged the adolescents to make a personal decision to abstain from using drugs based on ethical or moral considerations. The DARE type programs focused on ways to strengthen an individual's intrapersonal functioning to forestall the involvement with drugs (self-esteem building, self-acceptance, feelings of competence), and also included some interpersonal skills to strengthen social functioning.

Developmentally, the intrapersonal focus with its goal of increasing self-esteem may have greater potential in the elementary grades. Elementary students are usually in contained classrooms or with a single teacher for most of the day, allowing the individual attention and recognition necessary for this type of approach. A junior high school teacher, in contrast, can be involved with upwards of 120 adolescents daily (four to five classes), which makes using this approach particularly difficult. Many leaders reported that the K+A programs were hard to implement (Hansen et al. 1988; Schaps et al. 1981).

Focus of Interactive Programs. Interactive programs focus primarily on interpersonal competence, and peer pressure is assumed to be the paramount reason for adolescents' use of drugs. Newcomb and Bentler (1989) identified peer influences as the "most consistent and strongest of all factors, influencing the ‘average' youth" (p. 245). Therefore, two types of peer pressure were central to the interactive programs. First, drug refusal skills were used in all SI and CLS programs to enable the adolescents to skillfully negotiate the refusal of a drug offer and simultaneously remain accepted by their peer group. Second, peer pressure can take another form: "[P]eer influence may result from perception of peer attitudes and behaviors rather than from actual peer behavior" (Beisecker 1991, p. 234). Krohn and colleagues (1982) found that adolescent drug behaviors were determined by the "norm qualities of friends (compared to parents and religion)" and this "is clearly the most predictive variable" (p. 343).

Adolescents usually overestimate the drug use of their friends and other peers. Normative education was used to challenge the adolescent's perceptions. Firsthand, through peer-to-peer interactions, adolescents learn about their acquaintances' drug use or lack thereof. Also, through the leader's input, information was provided about the local school, State, and national levels of adolescent drug use. The presumption was that, as adolescents develop more realistic perceptions, their anxieties related to peer pressure will be reduced and, in turn, their drug use. Although not used as frequently as drug refusal skills, normative education was a component in the majority of the SI and the CLS programs.

Adolescence is a period in which establishing peer relationships takes priority over adult relationships. The peer pressure issues central to the interactive programs are in contrast to the noninteractive programs, which depend upon an ethical decision or personal change of values. It comes as no surprise that the

interactive programs, based on peer-to-peer exchanges, were developmentally more appropriate and therefore more effective.

Interrelated Factors: Content, Process, and Size

The type of program was determined from a detailed coding for the nature of the program content and the type of delivery method or group process. The size of the program was identified as an influential third factor in this meta-analysis. Therefore, there exists a three-way interplay between the program content, the group process, and the size of the program. It is important to keep in mind that the focus of the content, in part, determines the method of delivery or group process. Until experimental studies compare identical content delivered with different group processes involving both small scale programs and larger programs, this puzzle will not be resolved. Notwithstanding, unequivocal statements can be made about certain combinations of content, process, and size that were identified in this collection of programs. To eliminate as many forms of bias as possible, only the values, the DARE type, SI, and the CLS programs included in the subset of 56 high-quality studies were used for the following comparisons.[18]

Noninteractive Programs. Comparisons cannot be made between the values programs and the DARE type programs, even though both were K+A programs and therefore used the same group process. The values (small K+A) programs were not implemented on a large scale and conversely, the DARE type (large K+A) programs were not implemented on a small scale. Comparisons cannot be made between the small unsuccessful values programs and the extremely successful SI and CLS programs because both the content and process differed. Only the following can be stated about the values programs: Content based on intrapersonal, ethical, and/or moral decisions that were presented in a noninteractive group did not change drug use behaviors, even when implemented in small programs.

Surprisingly, the large DARE type (K+A) programs appear to be somewhat more effective than the small values (K+A) programs. Statistical testing was not pursued because the two programs were not implemented in similarly sized programs. However, it would be expected that the larger DARE type programs would not do as well as the smaller values programs based on their size alone. Therefore, these results suggest that the values programs were particularly ineffective.

The DARE type programs were implemented in the same size programs as the large SI programs, but, again, comparison of the content or process cannot be made as both were different. What can be stated is that the combination of content and process used in the large SI programs showed statistically significant superiority when compared to the combination of content and process used by the DARE type (large K+A) programs. However, the content of the DARE type programs did show similarities to the CLS programs, although the group process[19] was different. Although both programs had an intrapersonal and interpersonal focus, the focus of the content in the DARE type programs was highly intrapersonal, with less emphasis on interpersonal drug skills. The CLS programs focused primarily on interpersonal skill building and to a lesser degree included a variety of intrapersonal components. The large CLS interactive programs were statistically superior to the large DARE type (K+A) noninteractive programs, indicating that the more interpersonal emphasis used in an interactive group was more effective.

Interactive Programs. Fortunately, the SI and CLS programs were represented as both small and large programs. Of equal importance, both the SI and CLS programs used an interactive group to deliver the content, and therefore the content can be compared. The SI programs stressed varied aspects of the social context that influences drug use and combined this with mastery of drug refusal skills. The more comprehensive CLS programs added many generic skills to the content of the SI programs and, in some cases, included both an intrapersonal and an interpersonal focus. Within the set of small programs, there were no statistically significant differences between the more singularly focused SI programs and the CLS programs, although the CLS programs had slightly higher effect sizes. For the large programs, the SI and CLS programs showed identical effectiveness.

Comprehensiveness of Content. The comprehensiveness of the program content appears to have no impact, as seen in the above comparison of the SI and CLS programs. Before controlling for the size of the program, it appeared that the more comprehensive CLS programs were more effective than the more singular SI programs. The possibility existed that inclusion of more content would increase effectiveness. However, closer examination of the nature of the additional content showed that its focus was expanded in some of the CLS programs to include both an interpersonal and intrapersonal focus (Botvin and Dusenbury 1989). Yet, other CLS programs

maintained the interpersonal focus even though using a more comprehensive content (Schnicke and Gilchrist 1984). In the smaller programs, the additional components did produce programs with somewhat higher effect sizes than the more singular SI programs, but this was not statistically significant. When comparing the large programs, the effect sizes for the CLS programs were equal to the large SI programs. For both the small and large programs, statistically significant evidence does not exist to support the greater efficacy of the more comprehensive CLS programs.

Five Remaining Covariates

Targeted Drug. The interactive programs were consistently much higher than the noninteractive whether the program targeted cigarettes or alcohol or did not target a specific drug. Only the WES for the noninteractive alcohol programs came close to the interactive alcohol programs. For the interactive programs, the generic drug prevention programs were at least three times more effective in preventing cigarette, alcohol, marijuana, and other drug use than the noninteractive programs.

Another question remains. How effective were programs which target a specific drug versus including that drug in a generic drug prevention program? There appears to be a considerable advantage in targeting cigarettes when using an interactive program, particularly if the program was implemented on a small scale. A possible explanation could be implementation problems experienced when small programs involved in efficacy trials were delivered under real world conditions. However, the evidence from the generic programs necessitates a different conjecture. Whether large or small, the generic programs within the interactive programs were approximately three times less effective than the smoking programs in preventing cigarette use. As size was not a factor, possibly the generic programs were less successful because an implicit message of lifetime abstinence was delivered in the smoking programs, while the generic programs, intentionally or unintentionally, may deliver a message of abstinence only until the drugs can be purchased legally.

Targeting alcohol within the noninteractive programs produces results similar to not targeting alcohol (i.e., generic drug programs). Targeting alcohol appeared to decrease program effectiveness for alcohol use when compared to generic programs in the set of 114 programs. However, the slightly lower effect sizes for the alcohol programs were not statistically

significant in the subset of 56 high-quality experimental programs.

Leaders. No single type of leader produced mean effect sizes that were statistically significant. Mental health specialists,[20] whose degree training involves the skills necessary to promote active group participation, were the most effective leaders, although not significantly so. However, only 29 percent of the mental health specialists were involved in large scale implementations associated with lower effect sizes. Peer leaders were used by only 21 percent of the programs. More often, the peer leader was a copartner with an adult leader. Peer leadership does not define an interactive program, nor was it a prerequisite for promoting the necessary group interaction, but peer leaders can be helpful in setting the stage and in supportive roles. In fact, it appears that the credentials of the leader may not be the issue as much as whether the leader can facilitate the necessary group interactions. The use of outside professionals may be questionable despite their level of skills, particularly if they are in the building only once a week for the drug prevention program. Also, a larger challenge remains. Can teachers create the atmosphere necessary for a truly interactive group when they have not been trained in the use of group skills, typically do not use the group process to present the course content, and must act as disciplinarians throughout the day?

Type of Control Group. Differential effectiveness was observed for the different types of control groups. The effect sizes were attenuated when the comparison/control group was a standard health class and/or another treatment. The differential between the two types of control groups was identical whether the programs were noninteractive or interactive. Programs that were compared to a no-treatment group reported an UNES about 0.08 higher than those compared to the standard health class. Not surprisingly, this difference was equivalent to the UNES of 0.09 for the KO programs (typical health classes). The two types of control groups appeared with equal frequency within the noninteractive or interactive programs; therefore, this variable does not contribute to the difference in effectiveness between the two types of programs.

Analyses of the type of control group, however, highlight another important issue for drug prevention program research. Since the 1986 Drug Free Schools and Community Act, few true no-treatment groups exist. In 1993, 38 percent of the programs were compared to the standard health class/another treatment control groups, an increase of 12 percent from previous findings

in 1986. This trend can be expected to continue; therefore, researchers should include information about the program content and the delivery method used by the standard health class control group. This information was seldom reported and is extremely important. For example, Ary and colleagues (1989) found that the control schools were receiving the same number of sessions (12) as the treatment schools. In actuality, Ary and colleagues' program was compared to another program of equal strength. In the same vein, it is the rare school system that does not have drug prevention activities (e.g., assemblies, drug prevention week). Therefore, evaluations should mention all other drug education activities and the amount of previous exposure to drug prevention programs.

Experimental Design. A pervasive drug prevention research problem is high dropout rates, that is, experimental mortality. This problem is exacerbated by the fact that drug users drop out of programs (even school-based) at higher rates. This was confirmed by 63 percent of the programs in this meta-analysis. In most cases, the dropouts come from highly transient populations. Frequent moves can be indicative of unstable families, which have been correlated to higher drug use rates among adolescents (Ellickson et al. 1988). If analyses indicate no differential dropout from the treatment group and the control group, the internal validity of the experimental design has not been compromised. Unfortunately, only 37 percent of the programs reported this information. When a program is successful with the drug-using population, the high attrition rates could restrict the magnitude of program effectiveness (i.e., users were not present to show decreased use). This meta-analysis showed higher mortality rates were indeed associated with slightly lower effect sizes in the interactive programs, although no differences were observed for the noninteractive programs.

Special Populations. Programs including minority student populations were equally or slightly more effective than those delivered to white populations across two situations: large programs were highly successful when the school had a combined minority population over 50 percent, and small programs designed to be culturally sensitive were equally successful with black, Hispanic, or Native American adolescents.

Intensity

The explanation of no significant findings for the intensity variable was possibly related to the low intensity of both the noninteractive and interactive programs; both were only 10 hours. Sixty-eight percent of all the programs included were low intensity programs with a mean delivery intensity of only 6 hours; only 16 programs offered boosters, and only 4 provided more than 1 year of boosters. Positive behavioral effects were found for the interactive programs at an intensity of 10 hours, in contrast to findings of a national survey of 4,738 students in grades 3 to 12 in which no behavioral effects were observed at the end of 1 year of health education, although positive behavioral effects did appear at the end of 3 years of continuous health education (Health Education Works 1990). Similarly, the School Health Education Evaluation found "'medium' effects are achievable for general health practices when more than 30 hours of classroom instruction is provided" (Connell et al. 1985, p. 321).

Levels of Use

Although evaluators are increasingly determining program success based on an individual's initial level of drug use, only 35.8 percent (43) of the programs had classified participants by level of drug use. For more than a decade, Goodstadt (1986) has advocated measuring program success based on a participant's prior level of drug use to determine if differential effectiveness existed. A priori, Dielman and colleagues (1990) separated program participants on previous drinking experience. Without separate analyses, Dielman and colleagues found that some effects would have been attenuated and others would have been completely masked. Biglan and colleagues (1987) also illustrated the danger in lumping all subpopulations together. In this case, the nonsignificant findings for the nonsmokers completely overshadowed the highly significant findings for smokers.

The interactive programs were generally successful with smokers, as evidenced by five small but highly effective program outliers. The program outliers were excluded from the regressions, therefore, they will be mentioned here. Two separate SI programs were implemented with impoverished inner-city black males. One program was highly successful with cigarette users (Spitzzeri and Jason 1979). The second found limited effects with cigarette users, but was highly successful with experimental cigarette smokers (Jason et al. 1982). The remaining three program outliers[21] were variations of the subcategory of interactive programs called "others" and were

implemented with high school cigarette smokers (Greenberg and Deputat 1978). Logically, these programs did not include refusal skills as these adolescents were smokers who were dealing with cessation issues. The content centered on knowledge of the physical effects and health risks associated with smoking. An age-appropriate, less structured interactive group (group D) was used to convey this information. These outliers suggest that it might be beneficial to separate out the cigarette smokers with a distinctly different type of program, particularly at high school age.

Implementation

Drops in the magnitude of effectiveness experienced by the large programs suggest that factors other than statistical leveling of effect sizes (observed as the number of participants increases) were operating. Implementation factors seem to be a more probable explanation and a crucial mediating factor in determining the success of a program. Was an essential ingredient of the interactive programs missing, that of active involvement and interaction between peers? Ideally, an interactive program incorporates participation by everyone, preferably in small groups. To become proficient in the use of drug refusal skills or other new interpersonal skills, each individual needs a sufficient opportunity to practice before an assumption can be made that the skills can be transferred to actual drug use situations. If implemented in a regular classroom without extra leaders, the likelihood of every adolescent interacting on a regular basis to make this possible remains questionable. Along the same line, Botvin and colleagues (1990) found that some teachers did not include all parts of the program equally, possibly because they felt uncomfortable with certain areas such as the role plays. Other teachers "may not have been convinced that the approach being tested was as effective as teaching factual information about drugs and the adverse consequences of use" (p. 27). Botvin recommended extensive training to convince the teachers of the merits of this type of program and to provide the requisite skills and confidence necessary for implementation of this type of curriculum.

Replication of Findings With a Set of 56 High-Quality Experimental Programs

The quasi-experimental (nonrandomly assigned) programs were eliminated to rule out the possibility of positive bias. Instead of obtaining lower effect sizes, the set of 56 experimental programs

had higher effect sizes and an even larger differential between the noninteractive and interactive programs. An alternative explanation for the higher effect sizes may be the more stringent selection criteria used. Nevertheless, even these high-quality programs had flaws in their evaluations or implementations that may have positively or negatively biased the program effect size. Some sources of systematic error were removed in the subset of 56 studies; the remaining flaws, it is hoped, were random. For example, history may have affected some studies while implementation factors may have presented problems in others; but, when enough programs are included, these flaws may be considered random error. "What is systematic error in an individual study may be random error in the context of a meta-analysis" (Shotland and Mark 1987, p. 86).

Perhaps the confusion reported in the literature arises, to a larger degree, from including programs whose success was attenuated or inflated for one or more reasons. To address this question, the entire set of programs, whether randomly or nonrandomly assigned, was subjected to the more stringent selection criteria. The end result was a set of 68 programs, 56 experimental and 12 quasi-experimental. In table 6 the effect sizes are given for the entire set of 114 programs, the set of 68 programs that excluded programs with problems which could bias their results, and the final set of 56 programs that excluded problematic programs and also were randomly assigned. Following removal of problematic programs, there was an increase of 0.07 in the difference in both UNES and WES between the noninteractive and interactive programs. On the second step, removal of the quasi-experimental programs, the effect size difference between the noninteractive and interactive programs increased by 0.04 for the UNES and 0.02 for the WES. Ruling out the other sources of bias was far more important (nearly twice as much) as whether a program was randomly assigned. Additionally, only 34 percent of all the experimental programs were eliminated for other problems. Whereas 67 percent of the quasi-experimental programs had additional problems, a disproportionate number were compromised for reasons other than the lack of random assignment. The inclusion of quasi-experimental programs attenuated the magnitude of success of the interactive programs and made the results more ambiguous. Similarly, Hedges and Olkin (1985) found "that variations in the outcomes of well-controlled studies are considerably easier to model than are variations in the outcomes of poorly controlled studies" (p. 14).

Success Across Drugs and Populations. The success of the interactive programs was equivalent across all types of

substances: cigarettes, alcohol, marijuana, and all other illicit drugs. This finding contradicts the reviews that have reported the success of drug prevention programs for cigarettes, yet have failed to report equivalent success for alcohol and other drugs (Botvin 1990; Flay 1985*b*; Moskowitz 1989). It

TABLE 6. *Mean difference between noninteractive (NI) and interactive (I) UNES and WES without problematic and nonrandomly assigned programs.*

	114 programs			68 without problems			56 experimental without problems		
	X_{NI}	X_I	Diff	X_{NI}	X_I	Diff	X_{NI}	X_I	Diff
n	44	70		21	47		18	38	
UNES	0.06	0.25	0.19	0.02	0.28	0.26	0.02	0.32	0.30
WES	0.08	0.16	0.08	0.04	0.17	0.15	0.04	0.21	0.17

is this author's conjecture that the majority of reviews arrived at these conclusions because of the limited number and variety of programs included in their reviews and, additionally, because all types of prevention programs were lumped together.

The findings of the 1993 meta-analysis were similar to earlier findings in 1986 which showed the peer programs had equal success across all types of drugs. (In 1986, both the SI and CLS programs were included in one category called peer programs.) The lower effect sizes in 1993 may be the result of all adolescents receiving some form of drug prevention information in the last decade (i.e., media, school assemblies, community activities).

Encouragingly, the interactive programs were successful in schools with predominately minority populations. This also repeats the 1986 findings where peer programs were found equally successful with minority or white populations.

Similar Conclusions across Multiple Statistical Analyses. Remarkably consistent patterns were observed across the numerous and varied statistical procedures. The results reported here were a reanalysis of an earlier report to resolve the problem of extremely complex regression analyses with four separate size groups (i.e., too few programs for the

number of parameters). Additionally, interpretations of the regression analyses were augmented with detailed descriptive statistics for the 114 programs and then separately for the noninteractive and interactive programs. Finally, statistics developed specifically for meta- analysis were used to further verify the results of the regression analyses (e.g., homogeneity of effect size, model specification).

Equivalent Success for Five Extremely Large Programs.

The effectiveness of programs implemented on a large scale can be diminished by control groups subject to factors such as mandated drug curriculum, unmotivated teachers, incomplete implementation, and perhaps most important, a limited amount of small group interaction; the list is almost unending. Fortunately, for the sake of comparison, five of the six largest programs in this meta-analysis were SI programs. These five large implementations had a mean sample size of 6,516 tested students and achieved an WES of 0.13. This effect size was equivalent to the effect size for the remaining 32 SI programs (WES = 0.12), which had a mean sample size of 924 tested students. This WES was accomplished in spite of operating under real world conditions. Also, four of the five large SI programs were compared to a standard health class control and may have achieved an effect size of 0.21 had they been compared to a no-treatment control group (i.e., difference of 0.08 between no-treatment and a standard health class control group). The consistency of these results, even though small, provides a robust finding (Flay 1985*a*). In other words, "Two 0.06 results are much stronger evidence against the null than one 0.05; and 10 p's of 0.10 are stronger evidence against the null than 5 p's of 0.05" (Rosenthal 1990, p. 133).

CHALLENGES

The identification of the types of programs that work generates more questions: Why are people still using those programs that don't work, particularly across whole States (Ennett 1993; Klitzner 1988)? Have efforts stopped short of the goal and not made the successful programs available to the general school population in a marketable form that can be placed in the hand of the teachers or principals? Even when educators are informed about recent research and would choose an interactive program, to the author's knowledge (with one exception[22]), program curriculums are not in a form that could be implemented with minimal effort. The schools have only one choice—the noninteractive programs.

A second challenge is whether policymakers can be convinced to shift to the more interactive approaches and do so quickly. With drug education mandated in most States, answers are needed to the following questions. What would be the cost of "changing horses in midstream"? Is the small effect size for these programs worth the cost to the taxpayers? What will the impact be on parents and communities that have enthusiastically and energetically supported a program, only to find the program has minimal or no effect on adolescent drug use? Imagine what could happen if this community enthusiasm were marshaled to support programs that have already demonstrated the ability to prevent, delay, or decrease drug use. Project STAR (Students Taught Awareness and Resistance) in Kansas City (Pentz et al. 1986) and the Minnesota Heart Health Program (Perry et al. 1989) combined community support with an efficacious school-based program. Both achieved a WES of 0.20, nearly double the effect sizes of similar types of school-based programs implemented on a large scale without community involvement.

Third, what is being done to address other antecedents of adolescent drug use besides peer pressure? School-based programs that are offered only once, most often in junior high school, cannot be considered a silver bullet to last throughout adolescence. School hours occupy only a small part of an adolescent's day; therefore, these programs cannot be expected to "counter the range of powerful forces that operate outside the walls of the classroom and school" (Goodstadt 1987, p. 31).

Finally, the paramount question for school boards and administrators is whether they will provide the necessary class time, the extra personnel, and the aggressive teacher training in the use of interactive group process skills. These efforts would restore the operative ingredient that may have been missing from the larger interactive programs: that of active involvement, an opportunity to exchange ideas and discover alternative perspectives, and sufficient practice time to assimilate the new interpersonal skills.

NOTES

1. The reanalysis also included a correction for overrepresentation of some programs in Tobler (1986). Only one effect size per program strategy was reported.

2. In Tobler (1992*a*), 18 nonorthogonal planned comparisons, the result of an extremely fine-tuned coding scheme, were tested with the full set of 114 programs and also for 3 subsets grouped by size. The number of programs in each of the three size groups was less than 40; therefore, these analyses were open to spurious findings and may have lacked power to detect significant findings. However, this was offset by verifying the results using a second regression procedure, weighted structural regression (WSR). WSR was developed to alleviate problems of numerous, correlated predictors and limited sample sizes faced by social scientists (Pruzek and Lepak 1992).

3. The 1986 meta-analysis used modality to refer to the specific type or strategy of a program.

4. A comparison of the 1986 and 1993 terminology as well as a detailed discussion of the content and the delivery process can be found in Tobler (1993).

5. Means and standard deviations were reported in only 10 percent of the studies in Tobler (1986).

6. High-risk youth is defined as an individual who is a school dropout; has become pregnant; is economically disadvantaged; is the child of a drug or alcohol abuser; is a victim of physical, sexual, or psychological abuse; has committed a violent or delinquent act; has experienced mental health problems; has attempted suicide; or has experienced long-term physical pain due to injury [Public Law 99-570, Sec. 4122 (b)(2)].

7. The drug use etiology for these populations necessitates multimodal and markedly different types of prevention programs (Bry 1982; Hawkins et al. 1987; Swisher and Hu 1983; Wall et al. 1981).

8. The two community studies were excluded as they offered a variety of additional support over the 4 years.

9. The content areas termed "extracurricular activities" and "others" occurred very infrequently and were subsequently dropped.

10. The term "mental health specialists" includes counselors, psychologists, psychiatrists, Ph.D.s or the equivalent in human services, or graduate level social workers.

11. Additional information about the increment to R^2, F change, and its significance for the independent variable as well as any covariates that reached significance in the OLS and WLS regressions for 114 programs and the subset of 56 experimental programs can be found in Tobler (1994).

12. Literature surveys were used only to locate the programs. In all cases, the original report was obtained for the meta-analysis.

13. This group of programs belongs in the strategy type that was referred to as "alternative programs" (MOD5) in the meta-analysis of 143 programs (Tobler 1986).

14. Six programs outliers were identified in the regression analyses and removed, which reduced the set of 120 to 114.

15. Regression procedures could not be used because the number of covariates (N = 6) was too large for the number of programs with cigarette outcome measures (N = 38).

16. "Efficacy trials provide tests of whether a technology, treatment, procedure, or program does more good than harm when delivered under optimum conditions" (Flay 1986, p. 451).

17. Effectiveness trials are defined as "Trials to determine the effectiveness of an efficacious and acceptable program under real-world conditions of delivery/implementation" (Flay 1986, p. 459).

18. The KO, AO, and others programs had three or fewer programs in their categories and were not included here.

19. These programs were implemented by police officers who delivered the content with lectures and/or officer-directed discussions which were seldom broken into small groups to provide the necessary interaction for a strong interpersonal focus.

20. Only 3 of the 20 mental health specialists delivered noninteractive programs.

21. As well as being highly positive outliers, the three programs targeted only cigarette smokers and were excluded for not meeting the selection criteria (i.e., a conservative assumption that regular smokers are addicted).

22. Contact Gilbert J. Botvin, Ph.D., Professor and Director, Institute for Prevention Research, New York Hospital Cornell Medical Center, Room KB 201, 411 East 69th Street, New York, NY 10021. Telephone (212) 746-1270.

REFERENCES

Ary, D.; Biglan, A.; Glasgow, R.; Zoref, L.; Black, C.; Ochs, L.; Severson, H.; Kelly, R.; Weissman, W.; Lichtenstein, E.; Brozovsky, P.; and Wirt, F. *School-Based Tobacco Use Prevention Programs: Comparing a Social-Influence Curriculum to "Standard-Care" Curricula.* 1989 Available from Dennis V. Ary, Ph.D., Oregon Research Institute, Willamette Street, Eugene, OR 97401.

Bangert-Drowns, R. Review of developments in meta-analytic method. *Psychol Bull* 99(3):388-399, 1986.

Beisecker, A. Interpersonal approaches to drug abuse prevention. In: Donohew, L.; Sypher, H.; and Bukoski, W., eds. *Persuasive Communication and Drug Abuse Prevention*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1991. pp. 229-238.

Biglan, A.; Glasgow, R.; Ary, D.; Thompson, R.; Severson, H.; Lichtenstein, E.; Weissman, W.; Faller, C.; and Gallison, C. How generalizable are the effects of smoking prevention program? Refusal skills training and parent messsages in a teacher-administered program. *J Behav Med* 10(6):613-628, 1987.

Botvin, G. Substance abuse prevention: Theory, practice and effectiveness. In: Tonry M., and Wilson J., eds. *Drugs and Crime, Crime and Justice.* Vol. No. 11. Chicago: University of Chicago Press, 1990. pp. 461-520.

Botvin, G., and Dusenbury, L. Substance abuse prevention and the promotion of competence. In: Bond, L., and Compas, B., eds. *Primary Prevention and Promotion in the Schools*. Newbury Park, CA: Sage Publications, 1989. pp. 146-178.

Botvin, G.; Baker, E.; Filazolla, A.; and Botvin, E. A cognitive behavioral approach to substance abuse prevention: One year follow up. *Addict Behav* 15:47-63, 1990.

Bry, B. Reducing the incidence of adolescent problems through preventive intervention: One- and five-year follow-up. *Am J Community Psychol* 10(3):265-275, 1982.

Cohen, J., and Cohen, P. *Applied Multiple Regression Correlation Analysis for the Behavioral Sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates, 1983.

Connell, D.; Turner, R.; and Mason, F. Summary of findings of the School Health Education Evaluation: Health promotion effectiveness, implementation, and costs. *J Sch Health* 55(8):316-321, 1985.

Dielman, T.; Shope, J.; and Butchart, A. "Differential Effectiveness of an Elementary School-Based Prevention Program." Paper presented at the University of California, San Diego Extension Seminar, *What Do We Know about School-Based Prevention Strategies? Alcohol, Tobacco and Other Drugs,* San Diego, CA, March 17-20, 1990.

Edwards, G. *Reaching Out*. Barrytown, NY:, Station Hill Press/PULSE, 1972.

Ellickson, P.; Bell, R.; Thomas, M.; Robyn, A.; and Zellman, G. *Designing and Implementing Project ALERT. A Smoking and Drug Prevention Experiment.* Santa Monica, CA: Rand Corporation, 1988.

Ennett, S. How effective is Project DARE? In: Montoya, C.; Ringwalt, C.; Ryan, B.; and Zimmerman, R., eds. *Evaluating School-Linked Prevention Strategies; Alcohol, Tobacco, and Other Drugs*. San Diego, CA: University of California Press, 1993. pp. 51-61.

Flay, B. Psychosocial approaches to smoking prevention: A review of findings. *Health Psychol* 4(5):449-488, 1985*a.*

Flay, B. What we know about the social influences approach to smoking prevention: Review and recommendations. In: Bell, C., and Battjes, R., eds. *Prevention Research: Deterring Drug Abuse among Children and Adolescents.* National Institute on Drug Abuse Research Monograph 63. DHHS Pub. No. (ADM)85-1334. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1985*b*. pp. 67-112.

Flay, B. Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. *Prev Med* 15:451-474, 1986.

Giaconia, R., and Hedges, L. Identifying features of effective open education. *Rev Educ Res* 52(4):579-602, 1982.

Gilchrist, L.; Schinke, S.; Trimble, J.; and Cvetkovich, G. Skills enhancement to prevent substance abuse among American Indian adolescents. *Int J Addict* 22 (9):869-879, 1987.

Glass, G.; McGaw, B.; and Smith, M. *Meta-Analysis in Social Research*. Beverly Hills, CA: Sage Publications, 1981.

Goodstadt, M. School based drug education in North America: What is wrong? What can be done? *J Sch Health* 56(7): 278-281, 1986.

Goodstadt, M. Prevention strategies for drug abuse. *Issues Science Technol* 3(2):28-35, 1987.

Goodstadt, M. School-based education research findings: What have we learned? What can be done? In: Rey, K.; Faegre, C.; and Lowery, P., eds. *Prevention Research Findings: 1988*. Office for Substance Abuse Prevention Prevention Monograph 3. DHHS Pub. No. (ADM)91-1615. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1990. pp. 33-45.

Goodstadt, M., and Sheppard, M. Three approaches to alcohol education. *J Stud Alcohol* 44(2):362-380, 1983.

Greenberg, J., and Deputat, Z. Smoking intervention: Comparing three methods in a high school setting. *J Sch Health* 48(1):498-502, 1978.

Hansen, W. School-based substance abuse prevention: A review of the state of the art in curriculum, 1980-1990. *Health Educ Res* 7(3):403-430, 1992.

Hansen, W.; Johnson, D.; Flay, B.; Graham, J.; and Sobel, J. Affective and social influences approaches to the prevention of multiple substance abuse among seventh grade students: Results from Project SMART. *Prev Med* 17:135-154, 1988.

Hansen, W.; Tobler, N.; and Graham, J. Attrition in substance abuse prevention research: A meta-analysis of 85 longitudinally followed cohorts. *Eval Rev* 14(6): 677-685, 1990.

Hawkins, D.; Lishner, D.; Jenson, J.; and Catalano, R. Delinquents and drugs: What the evidence suggests about prevention and treatment programming. In: Brown, B., and Mills, A., eds. *Youth at High Risk for Substance Abuse.* Alcohol, Drug Abuse and Mental Health Administration. DHHS Pub. No. (ADM)87-1537. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1987. pp. 81-131.

Health education works. *Prev Forum*, January 17, 1990.

Hedges, L. Advances in statistical methods for meta-analysis. In: Cordray, D., and Lipsey, M., eds. *Evaluation and Statistical Review Annual.* Vol. 11. Beverly Hills, CA: Sage, 1986. pp. 731-748.

Hedges, L., and Olkin, I. *Statistical Methods for Meta-analysis.* New York: Academic Press, 1985.

Hunter, J., and Schmidt, F. *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings.* Newbury Park, CA: Sage Publications, 1990.

Jason, L.; Mollica, M.; and Ferrone, L. Evaluating an early secondary smoking prevention intervention. *Prev Med* 11(1):96-102, 1982.

Johnston, L.; Bachman, J.; and O'Malley, P. *Drug Use, Drinking, and Smoking: National Survey Results from High School, College, and Young Adult Populations, 1975-1988.* DHHS Pub. No. (ADM)89-1638. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1989.

Johnston, L.; O'Malley, P.; and Bachman, J. *Drug Use among American High School Students, College Students, and Other Young Adults: National Trends 1985*. DHHS Pub. No. (ADM)86-1450. National Institute on Drug Abuse. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1986.

Klitzner, M. *Report to Congress on the Nature and Effectiveness of Federal, State, and Local Drug Prevention/Education Programs, Part 2; An Assessment of the Research on School-Based Prevention Programs.* U.S. Department of Education, Office of Planning, Budget and Evaluation. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1988. pp. 1-47.

Krohn, M.; Akers, R.; Radosevich, M.; and Lanza-Kaduce, L. Norm qualities and adolescent drinking and drug behavior: The effects of norm quality and reference group on using and abusing alcohol and marijuana. *J Drug Issues* 12(4):343-360, 1982.

Light, R., and Pillemer, D. *Summing Up: The Science of Reviewing Research*. Cambridge, MA: Harvard University Press, 1984.

Lipsey, M. Juvenile delinquency treatment: A meta-analytic inquiry into the variability of effects. In: Cook, T.; Cooper, H.; Cordray, D.; Hartmann, H.; Hedges, L.; Light, R.; Louis, T.; and Mosteller, F., eds. *Meta-Analysis for Explanation: A Casebook.* New York: Russell Sage Foundation, 1992. pp. 83-128.

Moskowitz, J.; Malvin, J.; Schaeffer, G.; and Schaps, E. Evaluation of an affective development teacher training program. *J Prim Prev* 4(3):150-161, 1984.

Moskowitz, J. The primary prevention of alcohol problems: A critical review of the research literature. *J Stud Alcohol* 50:54-88, 1989.

Murray, D.; O'Connell, C.; Schmid, L.; and Perry, C. The validity of smoking self-reports by adolescents: A re-examination of the bogus pipeline procedures. *Addict Behav* 12:7-15, 1987.

Newcomb, M., and Bentler, P. Substance use and abuse among children and teenagers. *Am Psychologist* 44(2):242-248, 1989.

Oetting, E., and Beauvais, F. Adolescent drug use: Finding of national and local surveys. *J Consult Clin Psychol* 58(4):385-394, 1990.

Office of Substance Abuse Prevention. *Prevention Plus II: Tools for Creating and Sustaining Drug-Free Communities.* Office of Substance Abuse Prevention. DHHS Pub. No.(ADM)89-1649. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1989.

O'Malley, P.; Bachman, J.; and Johnston, L. Reliability and consistency in self-reports in drug use. *Int J Addict* 18(6):805-824, 1983.

Pechacek, T.; Murray, D.; Luepker, R.; Mittlemark, M.; Johnson, C.; and Shutz, J. Measurement of adolescent smoking behavior: Rationale and methods. *J Behav Med* 7(1):123-140, 1984.

Pentz, M.; Cormack, C.; Flay, B.; Hansen, W.; and Johnson, C.A. Balancing program and research integrity in community drug abuse prevention: Project STAR approach. *J Sch Health* 56(9):389-393, 1986.

Perry, C.; Klepp, K.; and Sillers, C. Community-wide strategies for cardiovascular health: The Minnesota Heart Health Program youth program. *Health Educ Res* 4(1):87-101, 1989.

Pirie, P.; Murray, D.; and Luepker, R. Smoking prevalence in a cohort of adolescents, including absentees, dropouts, and transfers. *Am J Public Health* 78(2):176-178, 1988.

Pruzek, R., and Lepak, G. Weighted structural regression: A broad class of adaptive methods for improving linear prediction. *Multivariate Behav Res* 27(1):95-129, 1992.

Rhodes, J., and Jason, L. The retrospective pretest: An alternative approach in evaluating drug prevention programs. *J Drug Educ* 17(4):345-356, 1987.

Ringwalt, C.; Curtain, T.; and Rosenblum, D. "A First-Year Evaluation of D.A.R.E. in Illinois." Paper presented at the University of California, San Diego Extension Seminar, *What Do We Know about School-Based Prevention Strategies? Alcohol, Tobacco, and Other Drugs.* San Diego, October 18-19, 1990.

Rosenthal, R. *Meta-Analytic Procedures for Social Research.* Applied Social Research Methods Series. Vol. 6. Beverly Hills, CA: Sage Publications, 1986.

Rosenthal, R. An evaluation of procedures and results. In: Wachter, K., and Straf, M., eds. *The Future of Meta-*

Analysis. New York: Russell Sage Foundation, 1990. pp. 123-133.

Schaps, E.; Moskowitz, J.; Condon, J.; and Malvin, J. *An Evaluation of an Innovative Drug Education Program: First Year Results.* Report prepared for the National Institute on Drug Abuse Prevention Branch, 1981.

Schinke, S., and Gilchrist, L. *Life Skills Counseling with Adolescents.* Baltimore: University Park Press, 1984.

Shotland, R., and Mark, M. Improving inferences from multiple methods. In: Mark, M., and Shotland, R., eds. *Multiple Methods in Program Evaluation. New Direction for Program Evaluation.* Vol. 35. San Francisco: Jossey-Bass Inc., 1987. pp. 77-94.

Single, E.; Kandel, D.; and Johnson, B. The reliability and validity of drug use responses in a large scale longitudinal survey. *J Drug Issues* 5:426-443, 1975.

Smith, M.; Glass, G.; and Miller, T. *Benefits of Psychotherapy*. Baltimore: Johns Hopkins University Press, 1980.

Spitzzeri, A., and Jason, L. Prevention and treatment of smoking in school age children. *J Drug Educ* 9(4):315-325, 1979.

SPSS Inc. *SPSS Reference Guide*. 1990. Available from SPSS., Inc., 444 North Michigan Avenue, Chicago, IL 60611.

Swisher, J., and Hu, T. Alternatives to drug abuse: Some are and some are not. In: Glynn, T.; Leukefeld, C.; and Ludford, J., eds. *Preventing Adolescent Drug Abuse: Intervention Strategies*. National Institute on Drug Abuse Research Monograph 47. DHHS Pub. No. (ADM)83-1820. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1983.

Tobler, N. *Measuring Drug Use Differences from Pretest to Posttest: A Probit Change Score*. 1985. Available from the author, Box 246, Sand Lake, NY 12153.

Tobler, N. Meta-analysis of 143 adolescent drug prevention programs: Quantitative outcomes results of program participants compared to a control or comparison group. *J Drug Issues* 16(4):537-567, 1986.

Tobler, N. *Meta-Analysis of Adolescent Drug Prevention Programs: Final Report*. Rockville, MD: National Institute on Drug Abuse, 1992*a.*

Tobler, N. Drug prevention programs can work: Research findings. *J Addict Dis* 11(3):1-28, 1992*b*.

Tobler, N. Updated meta-analysis of adolescent drug prevention programs. In: Montoya, C.; Ringwalt, C.; Ryan, B.; and Zimmerman, R., eds. *Evaluating School-linked*

Prevention Strategies: Alcohol, Tobacco, and Other Drugs. San Diego, CA: UCSD Extension, University of California, 1993. pp. 71-86.

Tobler, N. Meta-analytical issues for prevention intervention research. In: Seitz, L., and Collins, L., eds. *Advances in Data Analysis for Prevention Intervention Research.* National Institute on Drug Abuse Research Monograph 142. NIH Pub. No. 94-3599. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1994. pp. 342-403.

Toseland, R., and Rivas, R. *Introduction to Group Work Practice.* New York: MacMillan, Inc., 1984.

United States Congress. House. *Anti-Drug Abuse Act.* Public Law 99-570, 99th Congress (H.R. 5484), October 27, 1986.

Wall, S.; Hawkins, J.; Lishner, D.; and Fraser, M. *Juvenile Delinquency Prevention. A Compendium of Thirty-Six Program Models*. National Institute of Juvenile Justice and Delinquency Prevention. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1981.

AUTHOR

Nancy S. Tobler, Ph.D.
Research Associate Professor
School of Social Welfare
University at Albany
State University of New York
Albany, NY 12150

Click here to go to page 69