DOCUMENT RESUME

ED 373 855 JC 940 523

AUTHOR Burley, Hansel
TITLE Persistence: A Meta-Analysis of College Developmental Studies Programs.
PUB DATE 94
NOTE 18p.; Paper presented at the Annual Forum of the Association for Institutional Research (34th, New Orleans, LA, May 29-June 1, 1994).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Persistence; Basic Skills; *College Preparation; College Students; Community Colleges; *Developmental Studies Programs; Dropout Research; Meta Analysis; Program Evaluation; Remedial Programs; *School Holding Power; *Student Attrition; Two Year Colleges; Withdrawal (Education)

ABSTRACT

To obtain a quantitative meta-analysis of studies addressing the effects of community college Developmental Studies Programs (DSPs) on students enrolled in higher education, an analysis was undertaken of a sample of studies culled from journal articles, dissertations, unpublished works found in ERIC, and published and unpublished proceedings of conferences. Results were measured by "effect size," a quantitative way of describing how well the average student who received intervention performed relative to the average student who did not receive the intervention. An effect size of zero, for example, indicated that a participating student did no better or worse than a non-participating student, while a positive effect size indicate that participants performed better than an average student. Results of the study included the following: (1) effect size estimates for developmental studies English composition ranged from -2.25 to 2.33; (2) effect sizes for developmental math ranged from .47 to 2.3; (3) persistence rates for reading ranged from -.32 to .13; (4) the one study of persistence rates for study skills courses that was reviewed had an effect size of -.22; (5) the majority of studies (66%) reported positive effects, indicating that DSPs improved retention rates; (6) programs with strong learning theoretical bases work better than programs that are watered-down versions of regular college classes; and (7) evidence indicates that well thought-out, self-paced, and competency-based programs keep students from dropping out. Contains 19 references. (MAB)

Persistence: A Meta-analysis of College Developmental Studies Programs

by

Hansel Burley

Institutional Research, McLennan Community College

Waco, TX

Paper presented at the Annual Forum of

The Association for Institutional Research

New Orleans, LA May 29 - June1, 1994

Persistence: A Meta-Analysis of the Effects of

Developmental Studies Programs

Study Objectives

It was the purpose of this study to integrate studies of college developmental programs to determine their effectiveness on student persistence. The studies chosen for this analysis examined only the effects of the total developmental studies program against a non-treatment comparison group. The studies selected were categorized a English composition, reading, or math.

Rapid Growth of Developmental Studies Programs and Rationale for Study

Do college developmental studies programs increase the chances of underprepared college students' success in college? This question is vital when one considers the increasing numbers of students entering these programs and the increasing proportion of college budgets that these programs absorb. Moore (1970) reported that 95% of all institutions of higher learning had some sort of remedial program. Also, nearly 100% of all community colleges had (and still have) these programs (Friedlander, 1980). These numbers represent phenomenal growth. A survey on remedial reading found that under 10% of the nations's colleges offered reading in 1960 (Bullock, Madden, & Mallery, 1990). By 1984, nearly 84% offered reading (Bullock, Madden, & Mallery, 1990). Also, by 1980, over 90% of the nation's colleges offered learning skills programs (Henry, 1986; Wright & Cahalan, 1985). Finally, between 1975 and 1980 remedial enrollment at four year institutions increased 72% while total student population increased only 7% (Chang, 1983). In one specific instance, one-third of Texas community colleges reported having 75% or more students enrolled in remedial mathematics courses (Grable, 1988).

The most cited research on developmental studies programs is survey research, notably the series of Cross studies and series of Roueche et al. studies. Notably, Roueche (1968) reported that 90% of students in remedial programs either failed or withdrew from the remedial programs. However, a Kulik, Kulik, and Shwalb (1983) meta-analysis of sixty evaluation studies (published between 1935 and 1979) of high risk college students reported modest positive effects for programs serving high-risk students. This analysis, however, included only nine studies that specifically addressed remedial or developmental studies, and only three of these reported positive effects on GPA. None of these studies included data from the 1980s.

Alfred and Lum (1988) noted that the research on developmental studies has been deficient in the examination of multiple effects of selected student and institutional variables on academic achievement. They called for research that included multiple institutions and examined the combined effects of student and institutional variables on academic achievement (Alfred & Lum, 1988). Innovative teaching strategies, new classroom management techniques, and new classroom technologies have generated a body of research (Schonberger, 1985) that needs to be examined, reviewed, and integrated.

Limitations of the Study

This study is limited by the following:

First, this study is limited to the studies collected by the researcher. Though the researcher attempted to gather the entire population of studies, such a population is realistically unknowable. The inability to obtain some studies because of unavailability or prohibitive costs further hampered collection of the population of studies. Finally, these studies did not indicate the subject matter taught.

Second, this study is limited by the accuracy of the reporting of the studies used.

Third, this study used only one coder, thus subjecting the study to possible bias in the coding of each selected study. To lessen this bias, three months after the coding, each observation was checked for accuracy.

This check was later followed by a reliability check of a sample of studies. This check showed no coding problems.

Definitions of both independent and dependent variables, subjects, instruments, and experimental designs varied from study to study. This coder used certain global generalizations to unify the studies as to definitions, meanings, and characteristics. For example, usually students are placed in DSPs because they failed to satisfy the particular requirements for entrance into a particular institution's regular curriculum. Finding some objective equalizing factor for the differing entrance requirements for all post secondary institutions included in the study would have been impossible. Therefore, any study of college programs designed to help students prepare for the regular curriculum was included in this meta-analysis. The reader should note that the differences between "precollege" students at the Air Force Academy may be very different from "precollege" students at an urban community college. Differences in institutions were noted in the coding; however, these results were not included in this study because of the small subgroup sizes that developed. These results will be included in a reanalysis of these data.

Fourth, this study used in its population both weak and strong reports of experiments. Weak studies tend to be flawed in some manner, thus lowering the validity and reliability of their reported outcomes. The rationale for including these studies is found in the Procedure portion of this paper.

Definitions

Developmental, remedial, special studies, compensatory, or opportunities programs. For the purposes of this study, any program designed to enhance the knowledge, skills, and attitudes of students so that they may meet entrance requirement and successfully complete a course of study at that institution were classed as developmental. Institutions of higher learning have designated these programs to provide instruction in basic reading, writing, and mathematics skills for those students who perform below college standards. A more global definition describes these programs as " . . . any program, course, or other activity for students lacking the necessary skills to perform college work required by the institution" (Cahalan, 1986).

Many variables that affect the underprepared student qualify this definition including: The kind of high school preparation, choice of college, level of college entrance standards, rigor of entry level courses, and availability of remedial courses. Varying institutional quality, mission, and definition of "remedial" can further complicate a researcher's quest for precise definitions (Cahalan, 1986). In most minds, "remedial" has come to mean corrective or curative education that attempts to help students learn something that was not previously learned or that has been forgotten (Clowes, 1980).

Persistence. *For the purposes of this project,* "persistence" is calculated as the ratio of end-of-term remaining students to beginning-of-term entering students. This is an end of treatment effect.

Methodology

Glass, McGaw, and Smith (1981) outline the characteristics of meta-analysis as follows: First, meta-analysis is quantitative. It uses numbers and statistical methods for organizing and extracting information from large masses of data that are nearly incomprehensible by other means. Second, meta-analysis does not prejudge research findings in terms of research quality. The findings of studies are not ignored a priori by imposing what may be arbitrary and non-empirical criteria of research quality. The influence of study quality on findings is regarded as an a posteriori question. Finally, meta-analysis integrates different studies so that general conclusions will emerge.

Procedure

This study is a quantitative analysis of a sample of studies that address the effects of DSPs on students enrolled in higher education. That sample of studies includes articles published in journals, dissertations, unpublished works found in ERIC, and published and unpublished proceedings of conferences, or yearbooks.

This section describes procedures for locating studies, quantifying study outcomes, and describing study features. The procedure was as follows:

1. The literature search. First, several indexes to documents and bibliographies were searched. The first and most fruitful database was the Educational Resources Information Center (ERIC) index. ERIC provides abstracts of published and unpublished studies, including abstracts of papers presented at conferences, progress reports of ongoing research studies, studies sponsored by federal research programs, and final reports of educational projects. The second database searched was Dissertations Abstracts International (DAI). DAI is an index of doctoral dissertations submitted from nearly 400 universities. The third was Psyclit. Psyclit is a database on CD-ROM that abstracts articles found in psychology journals. The fourth was Sociofile, a CD-ROM database that abstracts sociology journals. Fifth, several databases were searched using DIALOG Information Services, an on-line database search service. These databases included Academic Index, (indexes education journals designed to reach practitioners), Social Sciences Citation Index, (indexes articles and lists the citations used in the articles), and Social Science Index, an index of articles in sociology journals. Finally, reference sections of studies selected yielded other studies not listed in any of the above databases.

The descriptors for the various computer searches were as follows: (1) "Higher education," (2) "Developmental," "remedial," "underprepared," "high-risk," "at-risk," (3) "composition," "reading," "mathematics" "special services," "study skills," and (4) many variations of "computer assisted instruction", "audio-tutorials," "auto-tutorials," "Personalized System of Instruction (PSI)" and "Keller Plan."

These searches yielded more than 3000 citations that addressed college DSPs. A reading of the abstracts of these studies or the studies themselves, revealed that only about 300 reported on quasi-experimental and experimental studies, and the pool was reduced accordingly. Copies of these studies were acquired. No predictive or correlational studies with control groups were found.

2. Selection of the studies. The studies were described, classified, and coded. The criteria for selecting the studies were borrowed from Kulik, Kulik, and Shwalb, (1983) and modified and enhanced for the requirements of this study. The criteria for selection of the studies were as follows: (1) studies must have

involved students enrolled in college DSPs, (2) studies must have reported on measured outcomes for groups of students in "special" programs or groups of students being taught using "innovative" teaching methods and must have reported on groups of students in comparison programs or in comparison groups receiving traditional instruction, (3) when several different articles described the same study, that study was counted only once, using the most complete report, (4) when a single article gave results for several semesters, only the results for the last semester were included, (5) when a single paper reported only the results for subgroups of students, these results were combined into a composite tally, (6) the study must not have made obvious reporting errors.

3. Coding of the studies. The point of measuring and coding study characteristics to correlate study findings with properties of the study. Since the goals of this study were quite broad, the coding needed to be comprehensive. Besides the recording of study outcomes, coding involved the recording of the "mediating effects" like instructional strategy or study quality.

Many of the studies reported multiple summary results. Only those deemed most important by the coder were entered into the data base. For example, a composition intervention may have reported results on three achievement dependent variables--a final writing sample grade, number of words produced on the final writing sample, and number of errors produced on the final writing sample. The researcher would have chosen the final writing sample grade. This decision is based on the heuristic that reported grades and test scores best describe a study's conclusion about a treatment.

The mediating effects, program characteristics, and summary results were first recorded in a data base program called Paradox. This computer program allowed the raw data to be recorded directly to a notebook computer via an on-screen form that looked like a paper coding sheet. The use of this computer program avoided the need to use perhaps over a thousand sheets of paper for coding. Paradox allowed for the sorting and manipulation the data in various ways. For example, in studies with multiple effect sizes, recopying program characteristics that did not change was an easy affair. This on-line coding sheet helped reduce noise in the data by skipping the steps of coding onto a sheet of paper, then entering the data into the computer. After recorded in Paradox and checked for accuracy, the data were transferred to a computer spreadsheet called QuattroPro which converted the data into American Standard Code for Information Interchange (ASCII), a format readable by SAS. QuattroPro also provided an excellent environment for making sure the spacing was correct and for checking the details of data set. For example, QuattroPro was used to insure that blank spaces that represented missing data were converted to dots.

Since this study used only one coder, what Glass, McGaw, and Smith (1981) called the "principal source of measurement unreliability" (p.75) (different coders not seeing or judging characteristics of a study in the same way) was not a problem. However, to ensure that the coding was reliable and that the report is stable, a reliability check was conducted. First, all of the studies used were checked for accurate coding. In the check, the coder compared each study to the printed output. Changes were made as necessary. Because of missing data in the study reports and failure to meet selection criteria, three studies and several observations from other studies were omitted. This check was followed by a formal reliability check of 10% of the studies. The results indicated no differences with the checked data. Therefore, it was concluded that the coding process was reliable.

4. Measuring study findings. The dependent variables for this statistical analysis were the outcomes of the developmental studies program, namely, achievement, attitude, or persistence. After the data were recorded, classified, and coded, the summary results were mathematically standardized, using effect size as the point estimate.

$$g=\frac{\overline{X_T}-\overline{X_C}}{s_*}$$

5. Analysis. The standard formula for estimating effect size was

where $\overline{X_T}$ is the treatment mean,

$\overline{X_C}$ is the control mean, and

$s_{\bullet}$ is either the control standard deviation or a pooled standard deviation (Hedges & Olkin, 1985).

Glass (1976) and Glass, McGaw, and Smith (1981) used the control standard deviation. It has the advantage of being uncontaminated by treatment effects (Hedges & Olkin, 1985; Hedges, Shymansky, & Woodworth, 1989).

However, Hedges and Olkin (1985) argued that the most precise estimate of the population variance is obtained by pooling, because the assumption of equal population variances is reasonable. Hedges and Olkin's (1985) modification of effect size is the same as above except in computing the standard deviation with the following formula

$$s=\sqrt{\frac{(n_T-1)(s_T)^2+(n_C-1)(s_C)^2}{n_T+n_C-2}}$$

Since (g) has a small sample bias, it is removed by using the (Hedges & Olkin, 1985) unbiased estimator of effect size

$$d\cong\left(1-\frac{3}{4N-9}\right)g,$$

where

$$N=n_T+n_C.$$

a. Both (g) and (d) (with pooled standard deviations) were calculated when both means and standard deviations were reported.

b. Some studies did not provide complete information. If possible, effect sizes were reported using alternate methods as discussed by Glass, McGaw, and Smith, (1981). For example, exact effect sizes were calculated as follows:

If (t) was reported and standard deviations are not reported then,

$$g=t\sqrt{\left(\frac{1}{n_T}+\frac{1}{n_C}\right)}.$$

However, this formula was not used for studies using matched pairs and where separate standard deviations are not reported (Hedges, Shymansky, & Woodworth, 1989).

If (F) was reported then,

$$\sqrt{F}=/t/.$$

Many other reported statistics were transformed into effect size estimates via a conversion of the reported statistic to a Pearson's (r) (Glass, McGaw, & Smith, 1981).

The conversion of Person's (r) to (t) is as follows:

$$t=r\sqrt{\frac{n-2}{1-r^2}}$$

as noted in Hinkle, Wiersma, and Jurs (1988).

c. Next, these effect sizes were then pooled (averaged) to create a common effect size

$$\bar{d}=\sum \frac{d}{K}$$

where K is the number of effect sizes.

(3) If the sample sizes differ from study to study, then the effect size estimate from large studies will be more precise than the estimates from smaller studies. This was controlled by weighting the more precise estimates when pooling (Hedges & Olkin. 1985). An approximate weight was given by

$$w_i \cong \frac{\tilde{n}_i}{\sum_{j=1}^{k} \tilde{n}_j},$$

$$\tilde{n} = \frac{n_T n_C}{n_T + n_C},$$

where the pooled weighted estimator is

$$d_w = w_1 d_1 + \ldots + w_k d_k .$$

(4) An ANOVA analogue (Hedges & Olkin, 1985) was performed to determine if all studies and the primary groups (reading, mathematics, composition, and total program) shared a common effect size for the dependent variables of achievement, persistence, and attitude. The formula for the homogeneity statistic (Q-total) is

$$Q_T = \sum_{i=1}^{k} \frac{d^2}{\sigma_d^2} - \frac{\left(\sum_{i=1}^{k} \frac{d}{\sigma_d^2} \right)^2}{\sum_{i=1}^{k} \frac{1}{\sigma_d^2}} .$$

The Hedges and Olkin (1985) process for determining the homogeneity of effect sizes is as follows (using the Chi-square table):

(a) If (Q-total) is not large or is statistically non-significant at the alpha .05 level, the process should be stopped, and it can be concluded that the single effect size fits the data adequately. (Note: This analysis adapted a generic SAS meta-analysis program provided by Hedges, Shymansky, and Woodworth (1989). In it they use the weighted mean effects in the SAS general linear model procedure to produce an (F) statistic. This statistic can be used just as (Q-

(b) A significant value of (Q-total) indicates that the effect sizes are not homogenous. The studies are then partitioned into groups by a particularly important dimension or characteristic. Since

$$Q_T = Q_B + Q_W$$

the within-class statistics was calculated.

If (Q-within) is nonsignificant, then for class (p) the class effect size was an estimate of the effect size of that group.

(c) A significant (Q-within) indicates that the effect sizes were not homogenous within classes. The values were examined to help identify classes with especially poor fit. This suggested other classifications to observe.

(d) Thus steps (b and c) are to be repeated until an acceptable level of within class fit is obtained or the possible classes are exhausted.

In accordance with the process noted above, (a) an overall (F) was computed. After that (b) the studies were divided into obvious and naturally occurring classes. For example, the studies were classified first by instructional methodology, then by subject content taught and instructional. The (F) for these subcategories was calculated. Those non-homogenous subcategories of the studies were divided into those with randomized designs and those without randomized designs. If non-randomized studies caused significance, they were removed from further study.

Two of the above analyses were done. The first used averaged effects within the study so that no matter how many effects sizes were coded, only one achievement, attitude, and/or persistence effect size from each study would be analyzed. This method protects the independence of the data.

Results

These studies, categorized as composition, mathematics, reading, and study skills) were categorized as studies of total programs. Effect size estimates for developmental studies English composition ranged from -2.25 to 2.33. Effect sizes for developmental math ranged from .47 to 2.3. Persistence rates for Reading ranged from -.32 to .13. The one study of persistence rates for study skills courses was -.22. See Table A.

Table A
Classification of Studies by Subject Taught and Instructional Strategy
Student Persistence

Author(s)	Instructional Strategy	N	ES	K	ES_k
English Composition					
Losak (1972)	Remedial Composition	433	-2.55	3	-0.77
McCormick & McCormick (1986)		435	-0.44		
Baranchick & Ladas (1979)		764	0.69		
Semke & Semke (1983)	Foreign language study	22	1.93		
Author(s)	Instructional Strategy	N	ES	K	ES_k
Benjamin, Kester, & Olson (1976)	Skills Center	1195	2		
Math					
Baranchik & Ladas (1979)	Remedial Math	939	0.47		
Benjamin, Kester, & Olson (1976)	Learning assistance center	973	2.3		
Reading					
Shea (1984)	Reading Program	100	-0.32	5	0.42
Baranchik & Ladas (1979)		1004	0.34		

Starks & Kuznik (1980)		250	0.67		
Gordon (1983)		96	0.7		
Maring, Shea, & Warner (1987)		80	0.74		
Francis (1984)	Counselling	48	0.13		

The majority of these studies (66%) reported positive effects.

Discussion and Conclusions

The statistic used for comparison purposes was effect size. Effect size is a quantitative way of describing how well the average student who received the intervention performed relative to the average student who did not receive the intervention (Borg & Gall, 1989). A 0 effect size means that an average student receiving the innovative instructional methodology did no better or worse that an average student not receiving the innovative methodology, while a positive effect size means that the average student receiving the intervention did better than the average student not receiving the intervention. For example, an effect size of .20 means that the average student in the experimental methodology scored one-fifth of a standard deviation higher than the average student in the control group. This would place that student at the 58th percentile in the control group. "Researchers, " according to Borg and Gall (1989), "consider effect sizes larger than .33 to have practical significance, that is, the effect is large enough to make a worthwhile difference in the outcome" (p. 7). However, Hinkle, Wiersma, and Jurs (1988), contrast Cohen's statement that an effect size of .25 was small with Feldt's conclusion that an effect size of .20 was small. They conclude, along with Cook and Campbell (cited in Hinkle, Wiersma, & Jurs, 1988), that consciousness of an appropriate effect size magnitude is a complex and perplexing issue. For the broad purposes of this study, determining practical significance of effect size involved considerations of the number of effects in the subcategory of studies, the nature of the subcategory itself, and mediating factors that may be in the studies. Therefore, effect sizes were compared directly--with caution. For instance, this researcher saw an instructional methodology requiring the latest technology, extra personnel, and extra staff development as having to have a higher mean effect size in order to have equal practical significance with programs, for example, that only require student scheduling changes. Practical significance requires an understanding of the fact that the context of the effect size is just as important as the magnitude of the effect size.

The data suggest that since Losak's (1972) study of a developmental studies program, that reported studies indicate improved retention rates. It would seem that those programs with strong learning theoretical bases work better than those programs that are watered-down versions of regular college classes. However, the effect sizes are based on effect size estimates, rather than exact effect sizes. Studies with large sample sizes produces larger than expected expected effect size estimates.

Still the evidence indicates that well though-out self-paced and competentcy-based programs keep students from dropping out.

References

Alfred, R. L., & Lum, G. D. (1988). Remedial program policies, student demographic characteristics, and academic achievement in community colleges. Community/Junior College Quarterly, 12, 107-120.

Borg, W. R., & Gall, M. D. (1989). Educational research: An introduction. (5th ed.) New York: Longman.

Bullock, T. L., Madden, D. A., & Mallery, A. L. (1990). Developmental education in American universities: Past, present, and future. Research & Teaching in Developmental Education, 6(2), 5-73.

Cahalan, M. (1986). College level remediation. (Fast Response Survey System Report No. 19). Rockville, MD: Westat Research. (ERIC Document Reproduction Service No. ED 274 290).

Chang, P. T. (1983). College developmental mathematics--a national survey. Paper presented at the Annual Convention of the Mathematical Association of America, Southeastern Section, Orlando, FL. (ERIC Document Reproduction Service No. ED 234 841).

Clowes, D. A. (1980). More than a definitional problem: Remedial, compensatory, and developmental education. Journal of Developmental and Remedial Education, 4(10), 8-10.

Cross, K. P. (1976). Accent on learning: Improving instruction and reshaping curriculum. San Francisco: Jossey-Bass.

Friedlander, J. (1980). Are college support programs and services reaching high-risk students? Journal of College Student Personnel, 21(1), 23-28.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). Meta-analysis in social research. Beverly Hills, CA: Sage.

Grable, J. R. (1988). Remedial education in Texas two-year colleges. Journal of Developmental Education, 12(2), 2-5.

Hedges, L. V., & Olkin, I. (1985). Statistical methods for meta-analysis. New York: Academic Press.

Hedges, L. V., Shymansky, J. A., & Woodworth, G. (1989). A practical guide to modern methods of meta-analysis. Washington, D.C.: Special Publications, National Science Teachers Association.

Henry, T. C. (1986). Needed: Comprehensive evaluation of education program efforts. Community College Review, 14(2), 46-52.

Kulik, C. C., Kulik, J. A., & Shwalb, B. J. (1983). College programs for high-risk and disadvantaged students: A meta-analysis of findings. Review of Educational Research, 53, 397-414.

Levine, A. (1978). Handbook on undergraduate curriculum. San Francisco: Jossey-Bass.

Moore, W. Jr. (1970). Against the odds. San Francisco: Jossey-Bass.

Roueche, J. E. (1968). Salvage, redirection or custody? Remedial education in the community college. Washington, D. C.: American Association of Community and Junior Colleges. (ERIC Document Reproduction Service No. ED 019 077).

Schonberger, A. K. (1985). Developmental mathematics in college: What the research is and why there isn't more. Paper presented at the research presession of the conference of the National Council of Teachers of Mathematics, San Antonio, TX. (ERIC Document Reproduction Service No. ED 256 439).

Wright, D. A., & Cahalan, M. W. (1985). Remedial/developmental studies in institutions of higher education policies and practices, 1984. (ERIC Document Reproduction Service No. ED 263 828).

Hansel Burley, Ph.D.
Institutional Effectiveness
McLennan Community College
1400 College
Waco, TX 76798
(817) 750-3425