EDUCATION LEADERS COUNCIL

ELC RESOLUTION ON ESEA REAUTHORIZATION

BEYOND ED-FLEX:

The Proper Federal/State Relationship

In The Delivery Of

K-12 Education

1001 CONNECTICUT AVENUE, NW • SUITE 204
WASHINGTON, D.C. 20036

TELEPHONE 202.822.9000

FACSIMILE 202.822.5077

The reauthorization of the Elementary and Secondary Education Act (ESEA) should be a time of bold innovation in American education. The federal role in K-12 education needs a fundamental overhaul, not incremental amending. The reform energies that abound in schools across America should be freed from federal red tape. The primacy of states in the provision of education must be recognized. Parents should be empowered. Local control should be respected. The dead hand of the federal bureaucracy and micro-management by thousands of pages of regulations attached to hundreds of separate programs should be ended. Three decades of ineffectual programs, wasting money, failing to accomplish their goals, and shackling innovation-minded local and state education leaders....all this needs to change. It's time for reform in Washington that's as fundamental and bold as the reforms now underway across America.

The members of the Education Leaders Council (ELC), who include the state education chiefs of Arizona, Colorado, Florida, Georgia, Michigan, Pennsylvania and Virginia, understand that education initiatives, policies and practices are most effective when generated closest to the children they aim to serve. Education policies fail when imposed upon communities by federal mandates and regulations. Congress can have a lasting impact in the delivery of education, by building on this principle in the reauthorization of ESEA.

All to often, federal, state and local education officials as well as many national education organizations focus on compliance with regulations rather than providing services that help children. Entrenched interest groups and bureaucratic fiefdoms, with their predilection for the status quo, have been hesitant about and even hostile to needed education reforms. Freedom to make decisions locally in exchange for accountability for results should guide federal education policy. What is needed today is a commitment to *trust* local educators, yet *verify* the accomplishment of children.

Pursuant to the "trust but verify" principle, ELC has called for unprecedented freedom for state innovation in exchange for the strong accountability measures which ESEA has always lacked. These are reflected in the following priorities:

- *Flexibility for states* in determining how federal dollars are spent and what they are spent on
- *Accountability* for results rather than simple compliance with spending categories, procedural controls and other regulations of input and process
- *High goals and expectations for all* students regardless of origin, income or location
- *Empowering parents, not bureaucracies,* by assuring that federal education dollars follow students to schools and programs of choice

The highly successful charter school movement sweeping through the states is a suitable model for a more productive partnership between the federal government

and the states. Simply put, states should receive wide-ranging freedom in the use of their federal dollars in exchange for significantly greater accountability for results.

As in a charter school agreement, a participating state (or large school district) would enter into a five year contract with the Secretary of Education in which it offers to produce specified academic improvements (and indicates how these will be measured). In exchange, participating states would gain broad flexibility in deciding how their federal education dollars are spent to achieve the agreed upon academic results. Sometimes termed "Super Ed-Flex", this strategy includes the following essential elements:

- A participating state enters into a five year contract with the Secretary of Education in which it agrees to produce specified academic gains and states precisely how these gains be measured
- States specify how federal dollars will be spent to attain agreed upon academic results
- Participating states choose which federal programs they want to include in "Super" Ed-Flex. Eligible programs include all formula-based K-12 programs, except IDEA
- Funding levels would be based on existing formulas. Monies from these programs may be commingled and spent as the state sees fit. All categorical program regulations are waived for participating states
- Accountability: contracts must include clear performance objectives and timetables for achieving academic improvement. Achievement must be disaggregated by student categories. If Title I is included in Super Ed Flex, achievement gains must be shown for disadvantaged children. If Bilingual Education funding is included, gains must be shown for LEP youngsters. As indicators, states may use State-level NAEP, a commercial test, state standards-based assessments or another mutually acceptable test of academic achievement
- Renewal: states that produce the specified results get their contracts renewed. States that fail to produce agreed upon results, revert to categorical/regulatory approach
- Participation: no state is required to participate and all have the option of continuing with categorical programs which remain on the books
- Evaluation: GAO evaluation is required prior to next ESEA reauthorization cycle

ELC members are confident that this approach will deepen and sustain the tide of freedom, innovation and accountability currently sweeping the landscape in the states. It goes far beyond the stale "block grant" debates of earlier years by linking greater freedom with improved results. We at the state and local level are focusing on standards with rigorous assessments, enterprise and accountability while pushing authority and control of curriculum and budgets to individual schools. Through innovations like charter schools, preoccupation with academic achievement, and renewed emphasis on performance, we are giving schools true autonomy with respect to budgeting, curriculum and personnel and meaningful choices to parents in exchange for accountability for results. We urge the Congress to do the same.

Chester E. Finn, Jr.
President

David H. Ponitz
Vice President

Thomas A. Holton
Secretary / Treasurer

THOMAS B. FORDHAM FOUNDATION
OUTSIDE THE BOX

Trustees
Chester E. Finn
Chester E. Finn, Jr.
Thomas A. Holton
Bruce Kovner
Bruno V. Manno
David H. Ponitz
Diane S. Ravitch

Embargoed for Release
Tuesday, February 23, 1999, 1 p.m.

Contact:
Shep Ranbom at (202) 467-8344

Federal Education Policy Must Take "New Directions," Leading Scholars and State Officials Urge

As Congress Works to Renew K-12 Programs, Experts and Policymakers Call Washington a Drag on Serious School Reform

WASHINGTON – February 23, 1999 – Federal programs aimed at improving public education have become an obstacle to significant school reform, according to a report released today by the Thomas B. Fordham Foundation in cooperation with the Manhattan Institute.

The findings are reported in *New Directions: Federal Education Policy in the Twenty-First Century,* a 175-page book edited by Marci Kanstoroom and Chester E. Finn, Jr. *New Directions*, which includes chapters by leading education scholars, journalists and state and local officials, offers bold ideas for reshaping federal education policy and urges Congress to consider fresh approaches in the upcoming reauthorization of the Elementary and Secondary Education Act (ESEA).

"Now is no time for tweaking," said Finn, president of the Foundation and John M. Olin Fellow at the Institute. "Thirty-four years and more than a hundred billion dollars after ESEA was enacted, it's clear that these programs have failed. They're out of touch, wrong-headed and ineffectual. They get in the way of serious education reform at the state and local levels. Congress should consider a major overhaul of the federal role in K-12 education. ESEA is the place to start."

In the book, education scholars Diane Ravitch and Paul T. Hill join Finn and others in blistering critiques of current federal policies and outline bold alternatives. Journalists appraise many of the present programs, while leading state and local officials, including Michigan Governor John Engler, explain how Washington is undermining their education reforms.

The book's release coincides with the winter meeting of the National Governors Association in Washington. Many governors are beginning to ask Congress to make major changes in how the federal government approaches primary and secondary education.

1627 K Street, NW • Suite 600 • Washington, DC 20006 • Telephone (202) 223-5452 • Fax (202) 223-9226
http://www.edexcellence.net • Publications: 1(888) 823-7474

"The federal education programs do worse than nothing," explains Governor Engler in the chapter he contributed to *New Directions*. "They divert and distract schools from their ultimate mission: educating children."

In *New Directions* Ravitch, a former assistant secretary of education and now a senior fellow at the Manhattan Institute and the Brookings Institution, calls for an overhaul of Title I, the largest of the federal education programs, by turning it into a "portable entitlement." She also urges that all federal aid follow children to the schools of their choice, that the present bilingual education program be converted into an "English literacy acquisition" program, and that states establish and enforce knowledge-based standards for teachers.

Of the present federal education programs, Ravitch writes, "All were established with high hopes, but none has lived up to the expectations of its sponsors. All are ripe for reform."

Hill, a professor at the University of Washington and senior fellow at the Brookings Institution, suggests consolidating all federal grant programs into a single funding stream and funneling the money to students based on need. He also calls for an expanded federal role in education research and development, and for a flexible fund to respond to pressing state and local education needs.

"No one can claim that the current chaotic system of laws, regulations, constraints, and preferences is the best one that we Americans could design for our children," writes Hill.

Finn also calls for turning Title I into a "portable entitlement" and funding it fully. He suggests the adoption of "Super Ed-Flex," which would allow reform-minded governors to consolidate federal education funds and spend them on their own reform priorities. In return, the states would be held strictly accountable for boosting academic achievement.

Lisa Graham Keegan, Arizona's superintendent of public instruction, agrees with this state-centered approach: "It is time to stop insisting that our representatives in Washington figure out what to do about education," she writes. "We'll figure it out for them."

###

Single copies of *New Directions: Federal Education Policy in the Twenty-First Century* can be obtained by calling 1-888-TBF-7474. The full volume is also available on the Thomas B. Fordham Foundation website: www.edexcellence.net.

The Thomas B. Fordham Foundation is a private foundation that supports research, publications, and action projects in elementary and secondary education reform at the national level and in the vicinity of Dayton, Ohio. It is not sponsored by, or affiliated with, Fordham University. *New Directions: Federal Education Policy in the Twenty-First Century* was published by the Foundation in cooperation with the Manhattan Institute, a New York-based research organization with a particular interest in urban policy.

A Federal Initiative for
CHARTER SCHOOL EXPANSION AND ENHANCEMENT

DRAFT for Discussion

Areas for potential federal assistance:

Charter Facilities Financing

Finding and financing a facility has proven one of the most daunting challenges facing most charter schools, particularly the undercapitalized community groups that have given the charter movement much of its promise and momentum. The federal government should back efforts to make it feasible for charter schools to obtain affordable, safe, and educationally appropriate facilities for their students. Specific proposals include:

- *Explicit inclusion of charter schools in any forthcoming general federal support for public school facilities*, such as the Qualified Zone Academy Bond program. As in other areas of federal policy, Congress and the U.S. Department of Education should ensure that charter schools have access to the same advantages other public schools enjoy.

- *Competitive funding for innovative approaches to charter school facilities financing*. The U.S. Department of Education would invite interested public and nonprofit organizations to submit proposals of innovative local approaches to addressing the charter school facilities challenge. Proposals that appeared most likely to help charter schools and to serve as national models would be funded and evaluated. Though no approaches would be prescribed, examples of initiatives could include:

    - Capitalizing low-interest loan pools for charter school capital needs.
    - Various credit-enhancement approaches such as loan guarantees and risk reserves.
    - "Pooling" initiatives, whereby some entity organizes schools into pools for the purposes of accessing lower-cost financing, such as tax-exempt bonds.
    - Programs of intensive real estate-oriented technical assistance at the state or local level.
    - Projects to develop "incubators" and other innovative real estate solutions for charter schools.

"Ripple Effects"

Assistance to Charter Authorizers

Supporting the Creation of "In-District" Charters or Charter Districts

A federal grant program, funded by consolidating existing ESEA resources, could be directed toward giving school districts incentives to:

- Support the creation of a significant number of **"in-district" charter schools** -- offering the flexibility-for-accountability tradeoff to district schools *independently* of state charter legislation. Schools would receive full per-pupil funding and broad operating autonomy in exchange for well-structured, clearly articulated accountability for results.
- **"Charterize" the entire district** (i.e., create a "charter district"), where: (1) all schools – both existing schools and new ones opened by teacher groups, community organizations, and other agencies -- operate on performance contracts and receive broad waivers from local and contractual regulations; and (2) district-wide school choice is available to all students. (Under this model, the functions and responsibilities of the school board and central office would shift from providing education to soliciting, managing, and monitoring an array of performance contracts designed to meet the community's diverse educational needs.)

"NEW CENTURY" SCHOOL DISTRICTS
A Federal Demonstration and Transitional Grant Program

DRAFT Discussion Paper

Introduction

The "New Century School Districts" initiative would offer local school districts incentives and resources to reinvent themselves for the 21st century. Through comprehensive school reform and charter schools, the nation has embarked on a bold effort to encourage citizens to recreate individual public schools from scratch. Large-scale reform is hindered, however, by the prevailing, inefficient paradigm of most school districts, under which the district owns and operates all public schools: controlling school budgeting and expenditures, as well as student and staffing assignments; regulating school structure and programming through district-level directives and constraints; and functioning as sole provider of goods and services needed by schools.

Just as government agencies around the world have reinvented themselves over the past two decades, there is great scope for reinventing the outmoded structure of our nation's school districts. Yet inertia at the local level makes it difficult to envision new approaches, and the transitional costs of moving to a new paradigm are often prohibitive for the tight budgets of most school districts. A federal demonstration program offering competitive grants could help overcome inertia, providing the inspiration and resources for pioneering districts to reinvent themselves.

Proposed Program

The U.S. Department of Education would invite applications from school districts of all sizes from across the country, proposing plans for bold and thorough redesign of the school district as an organizational and governance system. Funded proposals would be:

(1) genuinely bold in scope, yet feasible;
(2) substantially designed to increase educational opportunities and raise achievement levels for disadvantaged students;
(3) appropriate as potential models for similar efforts in other districts.

The program would offer preference to proposals from districts serving a substantial proportion of low-income students. Winning districts would receive substantial grants that could be used for a range of transitional needs, including hiring consultants or staff to design and implement a transitional plan. The demonstration grant would be transitional in nature [3-4 years??], as resource reallocation under the new paradigm would be expected to obviate the need for long-term federal funding. At the same time, continued funding throughout the term of the grant would be contingent on annual demonstration by districts of adequate progress toward meeting stated goals for raising student achievement (articulated in each district's proposal).

"New Century" districts would be required to participate in federal studies for the purpose of identifying and promoting successful models of "reinvented school districts" throughout the country.

Reinvention plans submitted by applicant districts might include (but need not be limited to) ideas such as the following:

- **Converting district service functions** to "enterprises" that operate on an exclusively fee-for-service basis.
- Creating and supporting **intra-district school choice** programs – including redesigning district planning processes regarding budgeting, capital improvements, enrollment levels, and personnel shifts to accommodate increased school choice.
- Constructing meaningful opportunities for parents, teachers, and community residents to **govern schools**.
- Supporting the creation of a significant number of **"in-district" charter schools** -- offering the flexibility-for-accountability tradeoff to district schools *independently* of state charter legislation. Schools would receive full per-pupil funding and broad operating autonomy in exchange for well-structured, clearly articulated accountability for results.
- "Charterizing" entire districts or creating **"charter districts,"** where: (1) all schools – both existing schools and new ones opened by teacher groups, community organizations, and other agencies -- operate on performance contracts and receive broad waivers from local and contractual regulations; and (2) district-wide school choice is available to all students. (Under this model, the functions and responsibilities of the school board and central office would shift from providing education to soliciting, managing, and monitoring an array of performance contracts designed to meet the community's diverse educational needs.)

The U.S. Department of Education would offer technical assistance to these efforts [is this helpful or ineffective [????], evaluate these initiatives over several years, disseminate lessons learned from the evaluations, and promote replication of successful models.

Funding for this demonstration program could be derived from consolidating and redirecting certain ESEA funding streams.

Making the Most of Summer School: A Meta-Analytic and Narrative Review

Harris Cooper
Kelly Charlton
Jeff C. Valentine
Laura Muhlenbruck
University of Missouri-Columbia

1/1/99
Draft manuscript, not for citation without author permission

Making the Most of Summer School: A Narrative and Meta-Analytic Review

Chapter 1
Introduction

Summertime, and the livin' is easy,
Fish are jumpin', and the cotton is high.
- From "Summertime" by George Gershwin (1935)

For most American children prior to 1935, Gershwin's lyrics did not capture the experience of summer as they knew it. In the 19th century, if children lived in agricultural areas, summertime meant helping tend crops or livestock around the farm. If children lived in urban areas, it was not unusual to attend school for at least two of summer's three months (Richmond, 1977). By the turn of the century, implementation of standardized school curricula and family mobility resulted in the current 9-month calendar, and summer became a time without school for children (Association of California School Administrators, 1988). Still, most children were immigrants from abroad who made their homes in large urban areas or they were part of the great migration of Americans from the farm to the city. For them, fish jumpin' and high cotton were things heard about in song, not experienced in everyday life.

If Gershwin's bucolic picture missed the mark, Hans Carste and Charles Tobias' ode to "those lazy, hazy, crazy days of summer" (recorded by Nat King Cole in 1962) was not much more accurate. As Doughtery (1981) pointed out, in the early 20th century, many children held jobs during the summer and those who were idle were a cause of concern for city dwellers. The passage of the first child labor law in 1916 meant that school-aged children had little, if anything, to do during their vacation from school. Community leaders demanded that organized recreational activities be made available for students when school was out. Education officials responded by creating the first summer schools.

Objectives of the Research Synthesis

Our primary goal in conducting this research synthesis was to gather, summarize, critically analyze, and integrate the research literature concerning the effects of summer school on children and adolescents. A good research synthesis should include both analytic and synthetic components. Analytically, it should point out and clarify, when possible, inconsistencies in the literature. One of our objectives, then, was to critically examine the research on summer school with an eye toward understanding why inconsistent results appear and how the methods chosen to study summer school might facilitate or restrict our ability to draw trustworthy inferences about program effects.

Synthetically, a good research synthesis should bring together what is known about a topic in the most parsimonious manner possible. In our case, this involved seeking an answer to the broadest question about summer school and also discovering the qualifications and caveats that restrict or specify that answer. Thus, we attempted to answer the question "What are the overall effects of summer school on children and adolescents?" We also sought qualifications to our response by answering the question "What characteristics of students, programs, and outcomes are associated with more or less effective summer programs?" The final characteristic of a good synthesis is that it should instigate a progressive paradigm shift (Strike & Posner,

Third, it is typically the case that the variables of substantive interest in a research synthesis (for example, whether summer programs are more effective for regular students or students with special needs) are confounded in study designs with methodological variations (for example, how carefully summer school participants and control students were equated). In this synthesis we will demonstrate a simple but rarely-used technique for adjusting effect sizes to remove the covariance between methodological and substantive characteristics within the meta-analytic data set.

Finally, we will show how to carry out a meta-analytic procedure that takes advantage of the fact that some studies contain within them variations of substantive interest to the synthesist. As our example, we found a set of studies that measured the effect of summer school on both reading and math. These studies provide for stronger causal inferences because most, if not all, other methodological, student, program, and outcome variations are held constant when effects on reading and math are compared. When multiple studies with within-study comparisons are available, meta-analysts can take advantage of them to draw stronger inferences.

History and Goals of Summer School

Today, the purposes of summer programs stretch far beyond making organized recreational activities available to students out of school. However, the *prevention of delinquent behavior* certainly remains among summer school's latent functions. Further, it is not unusual to find summer programs with a component that attempts to *build positive attitudes toward self and learning* to lessen the chances of anti-social behaviors in the future.

While keeping kids off the street provided the initial impetus for summer school, by the 1950s educators realized that summer schools could furnish opportunities to *remediate or prevent learning deficits* (Austin, Rogers, & Walbesser, 1972). Because the wealthy were able to hire tutors for their children, the educational summer programs made available through schools largely served students from disadvantaged backgrounds.

Summer programs to remediate learning deficits can be grouped into four categories. First, some summer programs are meant to help students meet minimum competency requirements for graduation or grade promotion. Most states set minimum competency standards for obtaining a high school diploma. When a student does not meet the standards, summer schools provide an excellent opportunity for the intensive instruction needed to raise the specific competencies above the required minimum.

Second, students who fail a particular course during the regular academic year use summer school as an opportunity to retake the course. Repetition of failed courses occurs primarily at the secondary level. The related remedial activity in elementary grades occurs when students are requested or required to take summer instruction in order to avoid being retained at the same grade level during the next school year.

Third, the movement to insure that students with disabilities receive a free and appropriate education (FAPE) resulted in the passage in 1975 of the Individuals with Disabilities Education Act (IDEA; Public Law 94-142). Prior to 1975, approximately 1 million children with disabilities were excluded entirely from public education, and more than half of all children with disabilities did not receive appropriate educational services (Federal Register, Vol. 62, No. 204, 1997). In 1979, the United States District Court ruled that the Pennsylvania Department of

programs take students with special talents to college during summer and provide in-residence programs.

Summer school also has been recognized as providing opportunities for teachers. Summer schools *offer teachers an opportunity to make additional money and to develop professional competencies*. While many teachers look toward summer as a time for relaxation, recreation, and rejuvenation, others use the opportunity to teach in summer as a way to supplement their income and to develop additional expertise.

Summer learning loss. A final rationale for summer school that has recently gained attention concerns the phenomenon of summer learning loss. A research synthesis conducted by Cooper, Nye, Charlton, Lindsay, and Greathouse (1996) helped kindle interest in this issue. These researchers integrated 39 studies examining the effects of summer vacation on achievement test scores. Their meta-analysis indicated that summer learning loss equaled at least one month of instruction. On average, children's tests scores were at least one month lower when they return to school in fall than scores were when students left in spring.

This meta-analysis also found dramatic differences in the effect of summer vacation on different skill areas. Summer loss was more pronounced for math than for reading and math facts and spelling showed much larger losses than other subject areas. To explain this finding, the authors suggested that both math facts and spelling skills involve the acquisition of factual and procedural knowledge whereas other skill areas, especially math concepts and problem-solving and reading comprehension, are more conceptually-based. Without practice, cognitive psychology suggests, facts and procedural skills are most susceptible to forgetting. The authors speculated that children's home environments provide more opportunities to practice reading skills than to practice mathematics.

In addition to the influence of subject area, the researchers had examined numerous differences among students as moderators of summer vacations' effect. Overall, the meta-analysis revealed little evidence to suggest that student intelligence had an impact on the effect of summer break. Also, neither the student's gender nor ethnicity appeared to have a consistent influence on the summer learning loss, although educators express special concern, likely legitimate, about the impact of summer vacation on the language skills of students who do not speak English at home.

Finally, family economics was examined as an influence on summer learning loss. The meta-analysis revealed that all students, regardless of economic group, lost roughly equal amounts of math skills over summer. However, substantial differences were found for reading. On some measures, middle class children showed gains in reading achievement over summer while disadvantaged children showed losses. Reading comprehension scores of both income groups declined, but more so for disadvantaged students. Thus, income differences appear to be related to differences in opportunities to practice and learn reading skills over summer; more books and reading opportunities are available to middle class children.

Cooper et.al. (1996) proposed two approaches to the alleviation of summer learning loss. First, the existence of summer learning loss could be used to argue for adopting changes in the school calendar (see Worsnop, 1996). Second, the summer loss suggested the continuation of summer remedial and enrichment programs. For all students, the authors suggested, a primary

thought to be the province of families. These would include teaching children positive character traits, physical or athletic skills, and how to maintain their health.

Maintenance of a globally-competitive education system. While new purposes for summer school will emerge, the primary focus is likely to remain academic. In the past two decades, many policy makers have become concerned about the global competitiveness of the American economy and the education system that drives it (National Commission on Excellence in Education, 1983). These concerns have focused considerable attention on the relatively short American school year. Writing in the Atlantic Monthly, Michael J. Barrett (1990), a former Massachusetts state senator and unsuccessful gubernatorial candidate in 1994, summarized this concern:

> The United States faces a time-in-school deficit every bit as serious as the trade deficit and the balance-of-payments problem: each year, American children receive hundreds of hours less schooling than many of their European and Asian mates, and the resulting harm promises to be cumulative and lasting. (p. 86)

Stevenson and Lee (1990), in a monograph entitled "Contexts of Achievement" that also appeared in this series, documented Barrett's concern. These researchers compared the number of days students spent in school in three countries. They found that the length of the school year was 230 days in Taipei, Taiwan, 243 days, in Sendai, China, and 174 days in Minneapolis, Minnesota. Students in Taipei and Sendai attended school for a half-day on Saturday. When Stevenson and Lee totaled the number of hours spent in school each year in the three countries, they found a roughly equal number of hours for 1st graders (because the Minneapolis school day was longer), but considerably fewer hours for Minneapolis 5th graders. While Minneapolis 5th graders spent about 1,044 hours in school each year, Sendai students spent about 1,466 hours and Taipei students spent about 1,655.

Also related to the need for summer school in the United States was Stevenson and Lee's finding that "American mothers may be dedicated to their child's development during his or her preschool years but ... they abdicate some of these responsibilities to the teacher once the child enters school. This trend is opposite from that which occurs in Chinese and Japanese families..." (p.100). Finally, students in the two Asian cities were found to spend more out-of-school time doing homework and reading than did Minneapolis students.

We should point out that Stevenson and Lee did not conclude that insufficient time and emphasis devoted to academic activities was the sole factor related to the Minneapolis students' relatively poor performance compared to their Asian counterparts, although it was first on their list (p. 103). They also indict inefficient use of time, low standards, and a lack of parent involvement, among others.

Establishment of high standards and minimum competency requirements. In addition to issues of global competitiveness, an emphasis has emerged nationally on higher academic standards and minimum competency requirements. The new standards and requirements have provided impetus for increased use of summer schools (Chmelynski, 1998). For example, Chicago Public Schools has a policy that establishes district-wide standards of promotion for students completing 3rd, 6th, and 8th grades. If students do not meet minimum grade-equivalent reading and math scores, report card grades, and attendance criteria, they are either retained or

more about intellectual development and cognitive growth by studying summer learning than by any other means" (p. III-15).

Ascher (1988) provided a brief review of research on summer schools, extended school years, and alternative school calendars for disadvantaged students. The review was prepared for the ERIC Clearinghouse on Urban Education. She concluded that attending summer school led to only modest improvements in elementary school students' achievement and at disproportionately high cost. Ascher also stated that "we know almost nothing about summer programs for disadvantaged middle and high school students" (p. 1). However, she noted that the poor results "can be at least partly attributed to the current programs themselves" (p.1). Problems with programs include short duration, loose organization, little time for advanced planning, low academic expectations, discontinuity between summer and regular school year curriculum, time wasted as new teachers get to know students, teacher fatigue, and poor attendance.

These three reviews are now seriously dated. More than a decade has passed since the last appeared. Further, none of the reviews claimed to be based on thorough searches of research sources. Finally, none used meta-analysis techniques to integrate the results of the included studies. All three reviews correctly suggested that even a thorough search followed by powerful synthesis tools would face numerous challenges because of the diversity among summer school programs and weaknesses in research. We turn our attention now to some of these challenges.

Issues in Summer School Research

Before examining the research in detail, it is important to set out some of the major issues researchers and research interpreters confront when attempting to evaluate the effectiveness of summer school. We can identify four interrelated issues that deserve special consideration.

Criteria for success. The variety of goals for summer school suggests that we will encounter equal variety when cataloging the criteria researchers have used to determine whether or not a particular program has been successful. Interestingly, we found no instance in which a measure of *delinquency* during the summer program was used as an indicator of program success. These measures might include archival data such as the arrest records of summer school participants. Occasionally, measures of the effects of summer school on factors related to delinquency, such as past participants' *absenteeism, self-esteem,* and *attitudes toward school*, have appeared in the literature.

With regard to *academic indices* of success, researchers have used criteria as simple as the percentage of students mastering learning objectives during the summer session or the percentage of students successfully completing a previously failed course. These percentages were compared to a specified percentage that was set by program overseers or evaluators. If the percentage was surpassed, the program was deemed a success.

The percentage-mastery criteria have the advantage of being both practical and easy to interpret. Their problem is that the specified percentage for declaring success can be set low, and at different benchmarks for different programs or for the same program in different years. This permits a host of subjective considerations to enter the evaluation process, sometimes predetermining the outcome. We found no evaluation of a summer program that used a percentage-mastery criteria that deemed the summer program a failure.

school year. Thus, a summer program might appear effective when, in fact, no learning gain had occurred.

For these reasons, the one-group pretest-posttest design provides an equivocal comparison for drawing inferences about program success. Its strength lies in our knowledge of the comparability of the pre-test and post-test groups, because they are the same students. Its weakness lies in the confound of the treatment with the passage of time between testings. We shall see that many evaluations of summer school employ this design, probably because of the ease of its implementation

Other designs used to assess the effects of summer school compare students who did and did not attend a program. In the best case, the *comparison groups are determined by randomly picking program participants* from a longer list of students who wished to go to summer school. Then, the researcher compares post-program scores for the picked and not-picked students.

A design of near-equal inferential power involves *identifying students who are eligible for summer school and then inviting a random subset.* Care must be taken in this design to track the number of invitees who refuse the offer. As the number of refusals increases so do differences between students who chose to attend and students not invited.

The use of random assignment in evaluations of summer school is rare, probably because denying access to a summer program on a random basis is politically untenable for schools and districts, unless more students want access than can be accommodated by the program. Therefore, less preferable methods are often used to constitute a group to compare to students who attend summer school. These methods involve using *matching or post hoc statistical procedures to enhance the comparability of program participants and nonparticipants.* Such procedures vary in rigor from study to study. For example, some investigators compare the post-program test scores of students who attend summer school with districtwide average test scores of students in the same grade. Others compare program participants with nonparticipants equated according to district, grade, achievement level, sex, ethnic group, and/or socioeconomic status.

The validity of these non-equivalent control group designs is a direct function of the care taken in the matching procedures. Obviously, the more numerous and relevant the equating variables, the more certain we can be that any post-program differences are actually due to summer school and not to pre-existing differences between students. However, our certainty using this design never reaches the level achieved by the use of random-assignment procedures (see Cook & Campbell, 1979, for a thorough discussion). The simple fact that summer school participants are motivated to attend the program often sets them apart from their peers in an important, perhaps critical, way. However, measuring motivation-to-learn often is difficult and costly. We found no instance in which researchers were able to equate non-randomly-assigned groups of participants and nonparticipants according to their level of motivation to do well in school.

Program persistence and attendance. Related to the issue of motivation is that of persistence among students who are in a summer program. Summer school is often optional and there are many enticements to drop-out or skip class. Therefore, program implementers often face high attrition and absence rates. Evaluators are confronted with the dilemma of whether to include in their data analyses students enrolled in the program who had poor attendance records.

the outcome measurements, the correspondence between goals and outcomes, the level of student participation, the achievement level and economic circumstances of participants, and, perhaps most critically, the nature of the comparison group. Moreover, these influences can have not only simple effects but also can interact with one another. We will illustrate this complexity by considering only a few variations on predictions for one subject (reading), and two groups of students, (elementary-school children from disadvantaged and middle class backgrounds).

We can predict that summer reading programs will have positive effects on disadvantaged children when program success is defined as the mastery of a small number of reading skills. If outcomes are measured by comparing student post-program scores to pre-program scores on standardized tests, we might expect that, for students from low-income backgrounds, the absence of a summer loss also indicates a program has been successful. However, a similar "no change" result for middle class children might suggest the reading program was not successful. The summer loss literature suggests positive change on selected reading abilities should occur for middle class students even in the absence of summer school. Predictions become more difficult if norm-referenced test scores are employed by evaluators. A gain of one month on grade-level equivalent scales essentially suggests that students have not changed their place in the distribution relative to the norming group. A change of more or less than one month is inherently ambiguous because it must be interpreted against (a) whether the test adjusted upward raw scores that revealed summer loss and (b) the length of the norming interval (typically, mid-fall and mid-spring) relative to the length of the summer testing interval.

Certainly, we can conclude a summer program is effective if randomly-assigned participants outperform their nonparticipating counterparts. We should expect an even larger effect of participation when summer school students are compared to nonparticipating students who have been matched on gender, economic, and/or achievement factors, but not on motivation. However, there is less reason to expect conclusions will be favorable to a summer program when comparing low-income students' post-program scores to districtwide average scores obtained at the same time. Again, summer loss research suggests that districtwide averages may include many nonparticipating students for whom reading improvements continue over summer, because they have greater economic support at home than program participants. This will be true even when achievement scores for both groups have been adjusted for initial achievement differences. We would expect the opposite relation to exist when middle class students who attend summer school are compared to districtwide averages. The middle class students then have both the program and their home environment working to their relative advantage.

We could continue to posit predictions. However, most of the multitude of permutations of interacting influences have no or few empirical testings described in the research literature. Rather than belabor the point, we now turn to a description and integration of research assessing the effectiveness of summer school programs.

searched for the period January, 1966 to August, 1998. The PsychINFO database was searched for the period January, 1967 to August, 1998. We used the search terms "summer school," "summer program," and "compensatory education" along with "summer" when searching both reference databases. After identifying 5111 documents through ERIC and 236 through PsychINFO we limited our search further to only those documents that referred to programs involving school-aged children. The search of ERIC identified 2065 document records that contained at least one of the search terms. The PsychINFO search identified 131 documents.

Two authors then reviewed the abstract of each record. We retrieved the complete document if either author believed the document might contain an empirical evaluation of the effects of summer school. This led to us review 258 full documents. Also, we examined the reference sections of reports that included empirical comparisons to determine if reference was made to other potentially relevant reports. We examined 75 reports found in other reference lists but not in the reference databases. A very few reports we would have liked to examine could not be obtained through our library or interlibrary loan.

Solicitation letters to active researchers in the field were not sent because we found no evidence that particular researchers pursued summer school evaluations on a continuing basis. However, we did receive several leads and unsolicited documents from researchers who were aware of our project.

Coding Frame

In many cases, characteristics of each study could be retrieved directly from the research reports with no inferences made by the coders. Information such as the sample size and the duration of the program was of this sort. In cases where some inference was necessary, we relied, first, on our own definitions of categories to code studies and, second, on definitions provided by the report writers, when no other information was available. For instance, with regard to a sample's economic background, if the sample was described as composed of children eligible to participate in a free lunch program or if data was given indicating that the average family in the sample earned income below the poverty line, we classified this sample as of "low socioeconomic status." If information about free lunch or family income was not given but the author said the students were drawn from a "poor" or "disadvantaged" neighborhood the sample was also classified as "low SES." Finally, we classified as low SES any program that was meant to serve children of migrant workers. Codes were left blank when reports did not contain enough information to make a confident judgment of a study's status on a characteristic.

Fifty-three different characteristics of each study were included in the database. The 53 characteristics could be classified as to whether they referred to aspects of (a) the research report, (b) the research design, (c) the sample of students, (d) the program's structure and content, or (e) the indicators of the outcome of the effectiveness of the summer program.

Report characteristics. Each database entry for a research report began with the name of the first author. Then, the type of report was coded as either a journal article (for published reports), dissertation, internal evaluation (typically a district level report), or an external evaluation (when an evaluation of a summer program was contained in a document issued by a private research firm or government organization other than the program sponsor). The year that the report was issued also was coded.

Program characteristics. The goals of the summer programs were categorized as "delinquency prevention", "promotion" after failure or retention, "remediation" of learning deficiencies among students who were below-average achievers, "prevention" of future academic problems, "improvement of attitudes" toward school, "academic enrichment" for all students, "development of underachievers," that is, students with above average ability not realizing their potential, "acceleration" of academic progress for above average achievers, or a "combination" of the above. Obviously, these labels are far from mutually exclusive and most often we found ourselves at the mercy of the categorizations used in the evaluation reports.

We coded the year in which the program was offered and how many years the program had been in existence. The size of the community served by the summer program was categorized as either a large urban area, a small city (with populations between 100,000 and 500,000), a suburb, a rural community, or a mixture of community types. The number of students, schools, and classes in the program was recorded, if given. The number of students in the program could differ from the sample size if not all students in the program participated in the evaluation.

We categorized programs with regard to whether participation was stated by the evaluators to be either required of the student or participants were invited or encouraged to attend. The average size of a class was coded as well as whether students were in residence at the summer school or traveled back and forth from home each day. We coded whether teachers were certified and whether the program required some involvement on the part of parents including, for example, volunteering time, observing a class, or taking part in a parent-teacher conference. We recorded data concerning the length of the program each day, and the number of days the program was in operation. Also, we noted the date of the last day the summer school program met.

The program's curriculum was categorized according to whether it was structured for "group" or "individual" instruction. The subject focus of the program was coded as math, reading, writing, language, science, other, or various combinations of subjects.

Outcome measures. We recorded whether outcome measures were taken immediately after the summer program ended, early in the next fall (September or October), late in the next fall (November or December), during the next winter or spring, after 1 year, or after more than one year. Also, we coded whether the outcome measures were teacher-constructed or standardized tests, and if standardized tests were used whether they involved raw test scores or some scaled measurement, such as grade-level equivalent scores.

A wide variety of knowledge and skill areas were measured as outcomes of the summer program. Specific subject matter outcomes included overall mathematics tests scores, as well as subtests on math concepts, computation, and applications, and overall reading scores, as well as reading comprehension, word and paragraph meaning, vocabulary, spelling, and word sounds. General language and communication skills and science knowledge also served as outcome measures. More general academic measures included IQ scores, Scholastic Aptitude Test scores, Cognitive Abilities Test scores, and class standing. In the nonacademic domain, several studies measured student self-concept and others measured self-help skills or attitudes toward specific subject matters.

Coder Reliability

After construction of the coding frame and its definitions, each relevant study was coded independently by two of the authors. Independent codes were then checked for consistency and disagreements were resolved in conference. Three authors regularly met during the coding and analysis process to discuss any codes that were done with low confidence. When coding was completed, most codes were rechecked before entry into the computer and during data analysis.

Search Outcomes

The literature search led us to 98 separate research reports that met the criteria for inclusion. However, two reports included the same data from a single program (Steinmiller, 1991, 1993) so only one report was included in the data set. Another report contained cumulative data on three years of a program whose first two years were covered in two other reports (Sipe, 1986, 1986, 1987). The first two reports were omitted from the data set. Two additional reports described summer programs and conducted empirical evaluations but did not report the information we needed to code a trustworthy direction for the effect of the program. Smith (1972) and the Texas Education Agency (1982) evaluated the effects of multiple Title I programs, including remedial summer programs, and concluded that their overall impact had been positive. However, neither evaluation reported data in a way that permitted us to separate the effect of the summer program from other programs. Thus, a total of 93 reports were coded for inclusion in our data set.

The 93 reports included evaluations of 89 separate summer school programs. Several reports described the outcomes of evaluations of the same program during different years of operation. Specifically, five programs had multiple reports describing evaluations conducted in different years (Petro, et.al., 1994, after Petro, et.al., 1993, after Rose, et.al., 1992; Donaldson, 1992, after Donaldson, 1990, after, Amorose, 1987; Robbins & Thompson, 1991, after Robbins & Thompson, 1989; Abram & Maurelli, 1980, after Abram & Cobb, 1979; Rawson, 1993, after Rawson, 1992). We coded these 12 reports as containing independent samples even though they referred to the same program. Once, the same summer program was described and evaluated in more than one report, with separate reports describing data collected on different samples of students during the same year (Siegelman, 1975, and Toledo, 1975).

Two reports included evaluations of two related summer programs during two consecutive years (Hanson, Yagi & Williams, 1986, after Kashmuk & Yagi, 1985). Two other reports included evaluations of three related programs during two consecutive years (Kulieke, 1986, and Olszewski, Kulieke & Willis, 1987). One report included data from one year of three separate summer programs (Kolloff & Moore, 1989). Another report (Klibanoff & Haggart, 1981) included evaluations of programs with two different goals.

Thirty-nine of the 93 reports included information on whether or not a summer program had a positive or negative effect on at least one academic-related outcome measure but did not include enough information to calculate d-indexes for the effect. Fifty-four reports included enough information to calculate effect sizes. First, we will examine the evaluations that provided only the direction of the effect.

comprehension. The authors reported that the standardized test scores were of "limited reliability" in evaluating the program, and that teachers reported improvements in skills, attitudes, and inter-group relations.

In a study by the Bureau of Evaluation in Louisiana (Rachel, 1987) the promotion and retention rates of 2nd through 5th grade students during the years 1980-1985 were examined. Data on summer school was collected at the state level for students participating in state-funded compensatory or remedial programs so very limited information was provided on the individual summer school programs. Pretest data was collected on 1623 2nd grade students during the spring of 1982, 4288 3rd grade students during the spring of 1983, 3729 4th grade students during the spring of 1984, and 6621 5th grade students during the spring of 1985. Thus, it is likely that there was overlap from year to year in the students who constituted the samples. The retention rates of those students attending summer school were compared to the retention rates of students not attending a summer school. At each grade level, a greater percentage of compensatory education students who had attended summer school were retained than among compensatory education students who had not attended a summer school. However, beyond their eligibility for compensatory education, no attempt was made to match students, so it is not clear what other factors (e.g., the severity of the learning deficiency) might have motivated a student's decision to attend summer school.

Curry et al. (1996) provided summary statistics for Title I summer programs conducted during the summer of 1995 in the Austin, Texas, school district. One-hundred-seventy-three kindergarten through 6th grade Title I students who had attended a summer program were compared with other Title I students. Summer school students in grades 5 and 6 showed higher standardized test scores in both reading and math than other Title I children, while students in grades two and three showed lower scores. Again however, no attempt was made to control for confounding factors that might have motivated some students to attend summer school and others not to attend.

Kushmuk and Yagi (1985) evaluated the 1983 and 1984 summer programs made available to students attending Portland, Oregon, schools. A variety of classes were provided including nonacademic (e.g., arts and crafts), and basic and remedial academic instruction. The authors did not break out results according to the type of class but did report separate results for students labeled as "not promoted," "enrichment," or "gifted and talented." Based on these labels we assumed that students who were not promoted attended summer school for either remedial or promotion goals. A pre-post design was used to compare the change in achievement test scores in math, reading, and language usage for 3rd through 7th grade student who attended summer school. No significance tests were conducted and no standard deviations were reported. District average gain scores were also provided but no attempt was made to create a matched control. Most groups of "not promoted" summer school students showed losses on reading test scores but gains on math and language scores.

Hanson, Yagi, and Williams (1986) evaluated the same summer program in 1985. That year, the board of education decided to waive tuition for students who qualified for free or reduced-price lunch. This time, separate results were reported for students labeled as "not promoted," "remedial," "enrichment," or "gifted and talented." The new policy nearly tripled the number of students attending summer school. Of 30 comparisons involving students who

number of independent samples that had all positive findings (n=95) divided by the number of findings that were either all positive or all negative findings (n=103). This proportion was .922. Using the mean condition sample size the estimated d-value was d=.12. Using the median sample size the estimated d-value was d=.49. Thus, we could say that using the vote-count procedure to provide a very rough estimate of the effect of remedial and promotion summer programs suggested that participants scored between one-tenth and one-half of a standard deviation higher on the outcome measure than their pre-program scores or the scores of nonparticpants.

Narrative descriptions of large evaluations. There were nine evaluations of remedial or promotion summer programs that provided only directional outcomes but were based on samples of more than 1000 students. We thought it might be informative to describe these studies.

An evaluation of fifteen migrant summer school programs in Colorado collected data on 2055 children in kindergarten through high school (Burnes, 1985). The author summarized data from both the 1983 and 1984 summer school. Approximately 40% of students received instruction in a language other than English. An attempt was also made to coordinate the delivery of other social services with the summer school program. Overall, positive gains in achievement test scores were made, with larger gains realized in math than in reading. Students in the 1983 summer program mastered fewer communication, math, English, and non-academic skills than did students in 1984.

Carroll (1987) evaluated three years of 25 summer programs throughout Connecticut. The programs were designed for children in kindergarten through 8th grade who needed remediation in reading, math, and language arts. Additionally, children were identified for participation if they were involved in a Chapter 1 program and had attendance problems. Achievement gains in reading, math, and language arts were found for approximately 540 kindergarten through 1st graders, 1304 2nd through 5th graders, and 396 6th through 8th graders.

In 1986, North Carolina began administering a program for testing 3rd, 6th, and 8th grader's basic skills (North Carolina, 1987). Those students who scored below a criterion on the North Carolina Minimum Skills Diagnostic Tests and students and who were retained in grade were invited to attend a basic education program in the summer. Two-thousand-eighty-seven 3rd graders, 4177 6th graders, and 5030 8th graders were given a basic skills tests before and after participation in the summer program. Students in all three grades in both 1986 and 1987 showed improvement.

The Utah State Office of Education evaluated Utah's migrant education programs conducted at 10 schools during the summer of 1988 (Moss, 1988). One of the ten programs was conducted during the regular school year as well as during the summer, the remaining 9 programs were limited to the summer. The length of the programs was varied, with some programs lasting 28 days and others lasting 50 days. In addition to academic instruction, students were provided with other enriching experiences, including a study of culture, recreation, and self-actualization, as well as basic health services. While 1,027 students participated in the various summer programs, evaluation data is provided for 404 students in grades kindergarten through 12. The Wide Range Achievement Test, Revised was given to students both prior to the summer program and after the program finished. Results indicated that the summer programs

obtain separate results by goal. In addition to the information given in Table 1, Table 2 also includes the goal(s) of the program.

The 10 reports included 45 separate independent samples generating 117 comparisons. Eighteen independent samples produced results that showed positive effects of summer school on all comparisons and 14 samples were predominantly positive. Five samples showed neither predominantly positive nor negative effects of summer school. Only one independent sample produced all negative effects and seven samples predominantly negative effects.

Narrative descriptions of predominantly negative evaluations. All of the independent samples that revealed negative effects of programs with other or multiple goals were contained in four reports.

Kashmuk and Yagi (1985), described above for their evaluation of Portland, Oregon, remedial summer programs, also presented data for students who attended summer school for enrichment purposes. Regular education children showed fairly consistent gains, although 4th graders declined in their reading scores and 6th and 7th graders declined in their language scores. Though gifted children showed consistent gains in reading and math, they showed consistent losses in language. The next year, Hanson et al. (1986) found that those students in summer school for enrichment gained most in pretest to posttest change in math while they lost most in reading and language. Similarly, gifted students showed gains in math and reading, but showed losses in language.

In Indianapolis, Indiana, the Marion County Public Library conducted a reading program that aimed to inspire children to a love of reading (Robbins & Thompson, 1991). Children who participated in the reading program were given control over what they read and how much they read. Incentives, such as ice cream, pencils, and fast food, were offered to children in an effort to encourage reading. The evaluation paid special attention to those children who were low achievers. One-hundred-sixty-four 1st through 6th grade students were evaluated for both vocabulary and reading comprehension. In general, students improved their vocabulary but student's reading comprehension scores declined from pretest to posttest. Despite losses in comprehension, the authors noted that "overall, students in the study maintained their reading achievement levels over the summer period" (Robbins & Thompson, 1991, pg. 19).

Seever (1991) evaluated the Kansas City, Missouri, summer school program that was implemented to provide remedial, developmental, and enrichment opportunities for kindergarten through 12th grade students. The summer school was conducted across five different sites, with one site offering a bilingual program. Seever concluded that there were "no sizeable differences between summer and spring testings" (1991, pg. 2). Indeed, an evaluation involving 865 6th, 7th, and 8th grade students showed a decline from pretest to posttest in both math and reading scores.

A vote-count estimate of effect size. Three of the 45 independent samples were missing data on sample size. Of the remaining 42, about 2,900 students participated in the evaluations. Thus, the average total sample size in each independent sample was about 70, and equal samples in treatment and control conditions would be 35. Again however, the data set contained a few evaluations with large samples, so the median total sample size was 35, or about 17 students in each condition.

Chapter 4

Meta-Analytic Procedures Used on Comparisons with Known Effect Sizes

Fifty-four evaluation reports contained enough information to permit us to calculate effect sizes measuring the impact of summer school on at least one outcome measure. These comparisons were synthesized using meta-analytic techniques. First, we will detail the statistical procedures and conventions used in the meta-analysis. A schematic representation of how the methods of meta-analysis proceeded can be found in Figure 1. Next, a brief statistical overview of the set of comparisons will be given. Finally, we will describe the substantive results of the meta-analysis. In the course of this description, each of the 54 studies in the database will be described, with special attention paid to their unique features or findings.

Methods of Meta-Analysis

Calculating average effect sizes. In our coding procedures, we described how we calculated the standardized mean difference, or d-index, between summer school participants and the control condition. Having these estimates in hand, we used both weighted and unweighted procedures to calculate average effect sizes across all comparisons.

In the unweighted procedure, each effect size was given equal weight in calculating the average d-value. We also identified the median d-value. In the weighted procedure, each independent effect size was first multiplied by the inverse of its variance and the sum of these products was then divided by the sum of the inverses. The weighting procedure gives greater weight to effect sizes based on larger samples and is the generally preferred alternative. Variance estimates for one-group pretest-posttest d-indexes were based on the formula $v = (1 + (d^2/2))/n$. Variance estimates for two-group d-indexes were calculated using formulas given in Cooper (1998). When we calculated d-indexes associated with categories of moderator variables, we only calculated weighted averages.

Also, 95% confidence intervals were calculated for weighted average d-indexes. If the confidence interval did not contain $d = 0$, then the null hypothesis of no summer school effect was rejected.

Identifying independent hypothesis tests. One problem that arises in calculating effect sizes involves deciding what constitutes an independent estimate of effect. Here, we used a shifting unit of analysis approach (Cooper, 1998). In this procedure, each effect size associated with a comparison is first coded as if it were an independent estimate of the treatment's impact. For example, if a single sample of students permitted comparisons of pre-program and post-program math and reading scores, two separate d-indexes were calculated. However, for estimating the overall effect of summer school, these two d-indexes were averaged prior to entry into the analysis, so that the sample only contributed one effect size. To calculate the overall weighted mean and confidence interval, this one effect size would be weighted by the inverse of its variance (based primarily on sample size, which should about be equal for the two component d-indexes). However, in an analysis that examined the effect of summer school on math and reading scores separately, this sample would contribute one effect size to each estimate of a category mean effect size.

It is rarely clear-cut which assumption, fixed or random, is most appropriate for a particular set of effect sizes. In practice, most meta-analysts opt for the fixed-effect assumption because it is analytically easier to manage. However, fixed-effects statistical models are often used when random effect models might be more appropriate. Indeed, a persuasive argument could be made that a random-effect model is most appropriate for the summer school literature. A random-effect model is indicated by (a) the large variation in summer program realizations and (b) our desire to draw inferences about all summer programs, not just those represented in our synthesis. Alternatively, it could be argued that a fixed-effect statistical model would be appropriate for our data if it is accompanied by a thorough search for moderators of effect sizes as part of the analytic strategy. Clearly, our strategy includes an exhaustive search for systematic moderating effects. Complicating matters further, evidence suggests that in the search for moderator variables, fixed-effect models seriously underestimate and random-effect models seriously overestimate error variance when their assumptions are violated (Overton, 1998).

Rather than opt for a single model that might be wrong, we chose to apply both models to our data. We conducted all our analyses twice, once employing fixed-effect assumptions and once random-effect assumptions. By employing this sensitivity analysis (Greenhouse & Iyengar, 1994), we could examine the effects of different assumptions on the outcomes of the synthesis. Differences in results based on which set of assumption was used could then be part of our interpretation of results.

Also with regard to interpretation, there are two fine points about the relationship between fixed-effect and random-effect analyses that should be stated before we begin to examine our results. First, on occasion the homogeneity analyses comparing groups of effect sizes using fixed-effect and random-effect models produce identical Q-statistics. This occurs when the study-level variance component equals 0. In the vast majority of instances, the study-level variance component is greater than 0, and therefore the variance of random-effect estimates is larger and Q-statistics comparing average effect sizes are smaller. This is why random-effect homogeneity analyses tend to be more conservative than the corresponding fixed-effect results.

Second, instances occur in which the weighted average d-value for fixed-effect and random-effect models are noticeably different. Recall that weighted average d-values give greater weight to individual effect sizes that are based on larger samples or, equivalently, have smaller variances. When a random-effect analysis is carried out a constant value, the study-level variance component, is added to each effect size's fixed-effect variance. Thus, the random-effect model tends to lessen the impact of sample size on the resulting weighted average d-value. This means that if the individual effect size magnitudes are related to sample size, the estimate of their average will change. When larger effect sizes are associated with smaller sample sizes the random-effect weighted average d-values will be larger than the corresponding fixed-effect estimate. When larger effect sizes are associated with larger sample sizes (a rare occurrence) the random-effect weighted average d-values will be larger than the corresponding fixed-effect estimate.

Statistical adjustment for methodological outliers. As is true for any data set composed of cases not under the control of the researcher, we confronted the possibility that significant correlation could exist among some of our moderator variables. For example, evaluations that used pretest-posttest change scores might also be more likely to sample students from economically-disadvantaged backgrounds. The confounding of methodological, student,

firm, 2 were evaluations conducted by university research centers, and one each appeared as a masters thesis, in a book, or as an ERIC document with no other identification.

Program goals. We found no program with the stated goal of preventing juvenile delinquency. Thirty-seven of the documents described programs that had as a goal the remediation of learning deficiencies. Three studies evaluated programs meant solely to prevent at-risk high school students from dropping-out (Curry, 1990; Wells, Springer & McCreedy, 1987; Opuni, et.al., 1990) or prevent "culturally-deprived" kindergartners from falling behind in first grade (Wasik & Sibley, 1969). Because of the low frequency of prevention program evaluations and because of the similarity in samples and content of these programs and remedial programs, we combined these two sets of evaluations together, bringing the total number of studies in the remediation category to 41 and the total number of independent samples in this category to 99.

The remaining 13 studies, accounting for 54 independent samples, were divided among programs with the goal of acceleration (7 studies, 34 samples), and programs with other or multiple goals (6 studies, 15 samples). One evaluation (Klibanoff & Haggart, 1981) included independent samples drawn from both remedial and multiple-goal (5 samples) summer programs.

Because of substantial differences in study characteristics, we decided to examine separately the evaluations of programs with the three different goals. The main meta-analysis relates to programs with the goal of remediation of learning deficiencies.

students "at-risk," 31 samples as "underachievers or below grade level," 26 samples as "failing or retained," and 18 samples as "learning disabled or otherwise impaired."

Programs evaluated in their first year of operation were compared to all programs evaluated after more than one year of operation. Samples drawn in large urban areas were compared to samples drawn from rural, suburban or small urban communities combined, with mixed and unknown community samples dropped from this analysis. Samples that were identified as comprised of volunteer students were compared to samples of nonvolunteers combined with samples for which we could not discern the sample's volunteer status.

About half (53%) of evaluations explicitly described the content of the summer curriculum and only three categories of content were used. Curricula were described as either specific to reading (9%), containing both a reading or language and a math component (16%), or including numerous, varied subject matters (27%).

About one-third of all posttest measures were administered immediately following the end of summer classes while another third were given in September or October of the year following summer school. About one-sixth had longer delays and these were combined into one category. About one-sixth of the reports did not mention when the posttest was given. The metric of the outcome measures were grouped into raw scores (85%) and all types of norm-referenced scales.

Finally, the coding of the subject matter of the outcome measures revealed about half of all measures were specific to reading and/or language arts and about 40% to mathematics. We found no measures specific to other subject areas. About 2% of measures related to general academic or intellectual abilities, and 6% to nonacademic areas such as attitudes toward school, self-esteem, and study habits.

Distribution and Measures of Central Tendency

Adjustment of statistical outliers and extreme values. First, we inspected the distribution of the 385 d-indexes using a stem-and-leaf plot. Applying Tukey's (1977) definition, we identified three positive d-index values that qualified as statistical outliers, that is, d-values that were more than three interquartile ranges beyond the 75th percentile. We discovered that these three d-values (d=4.40, 2.75, 2.71) came from two studies.

The single largest effect size came from an evaluation of a video-assisted intervention meant to improve the writing skills of students who had failed at least two courses during the previous school year (Welch & Jensen, 1990). Teachers were instructed in how to support the videos and implementation of the instructional model was monitored throughout a six-week session. The evaluators used both pre-treatment versus-post-treatment comparisons and posttest comparisons of treated and control students to assess the effect of the program. Both the students' knowledge of paragraph structure and student writing samples served as outcome measures. The pre-to-post comparison on the knowledge test revealed a d-index of 4.40. The report suggested that the content of the videos and the knowledge test were very closely aligned.

The second and third largest d-values came from a single evaluation. Cramer and Dorsey (1969) published a report on a summer reading program for poor-performing students in grades 4

basic academic skills and life skills instruction, and work experiences for two consecutive summers. Its three components included 90 hours of remedial instruction in reading and math, 18 hours of instruction in responsible social and sexual attitudes and behavior, and at least 80 hours of part-time work provided by the federally-funded Summer Youth Employment and Training Program. Support for the model development and pilot-testing were provided by numerous private foundations and public agencies. The programs ran in five urban areas: Boston, Fresno, Portland, San Diego, and Seattle. Both pretest-posttest and participant-versus-nonparticipant comparisons were provided for samples that ranged in size from 877 to 1876. The authors reported large learning losses over summer for control group students and lesser losses for program participants. However, only rough p-levels (i.e., $p < .10$, .05, and .01) and differences in means were provided without standard deviations. Therefore, d-values had to be estimated using sample sizes and p-levels. All the summer program effects were positive for treatment versus control comparisons, but some were not significant. Thus, the estimated values ranged from d=.01 the d=.15.

In addition to the six unusually large samples contained in the above three studies, we decided to Winsorize 14 other sample sizes greater than 500. We set these values to 500, and also set all treatment and control sample sizes (when these were known) to a maximum value of 250. We adopted these procedures because sample sizes are used to calculate the weights applied to each effect size when weighted averages are calculated. By restricting the weightings in this manner we ensured that no single sample could have an exceptionally large influence on the outcome of an analysis. In addition to the Seigelman (1975), Toledo (1975), and Sipe, Grossman, and Millner (1987) studies, three other evaluations included samples whose size was adjusted (D'Agnostino and Hiestand, 1995; Fonzi, 1984; Woloshin, 1975).

In the context of sample size, we should also mention an evaluation reported by Klibanoff and Haggart (1981), even though the individual sample sizes did not undergo adjustment. This evaluation was part of the large-scale study of the sustaining effects of compensatory education on basic skills. Data were collected on as many as 120,000 students in a nationally representative sample of elementary schools. Data on pre-program and post-program reading and math scores were presented for fifteen cohorts of students, ten in remedial summer programs, attending 1st through 5th grade. Sample sizes ranged from 53 to 190. Because this evaluation encompassed many schools and therefore many programs, little specific information was provided on the content of the summer programs. The sample effect sizes in this evaluation ranged from d=-.07 to d=.24.

Measures of central tendency and dispersion (see Figure 1). After Winsorizing the effect sizes, the average unweighted d-value for the 385 comparisons was d=.28. Using the 99 independent samples as the unit of analysis the average unweighted d-value was d=.32 The median effect size was d=.19. When the 99 independent effect sizes were weighted by the inverse of their variance the average d-index was d=.26. Thus, *making no distinctions among effects based on research methodology or program, student, or outcome characteristics, the average student outcome score after remedial summer school was over one-quarter of a standard deviation higher than the average student score before summer school or without summer school.* The 95% confidence interval for the weighted d-index under the assumption of a fixed

Tests for Methodological Moderators of Evaluation Outcomes

Table 4 presents the fixed-effect and random-effect Q-statistic along with the mean d-values and the variation around the mean d-value that encompasses the 95% confidence interval for the different levels of each methodological moderator. For example, if an average "Fixed d" equals .20 and its associated "+/-ci" equals .05 it means that the 95% confidence interval for the average value is .15 to .25. The reader might also note that if the d-value in the table is larger than the "+/-ci" then the d-value is significantly different from d=0 at $p<.05$.

Type of control group. The first moderator we examined concerned the type of control group used in the evaluation. First, we compared the size of the d-indexes in samples that used as the comparison pretest scores of students who attended summer school versus samples that used posttest scores from either nonequivalent groups or randomly-assigned groups of students who did not attend summer school. The fixed-effect homogeneity analyses revealed that the type of control comparison was a strong predictor of effect size, $Q(1,k=125) = 35.93$, $p<.0001$. The average d-index for one-group pretest-posttest designs was d=.24 while for two-group designs d=.07. (The number of samples exceeds 99 because a single independent sample could contribute more than one data point to this analysis if it contained both types of comparisons. This was also possible for a few other moderators described below). The random-effects analysis also found the type of control group to be significantly associated with variance in d-values, $Q(1,k=125) = 8.74$, $p<.005$. The random-effects average d-index for one-group pretest-posttest designs was d=.30 while for two-group designs d=.09.

Derivation of d-value. Within the pretest-posttest designs we next compared d-values that were derived from means and standard deviations with d-values derived from dependent t-tests. The fixed-effect model revealed significant differences between the derivation strategies, $Q(1, k=81) = 84.91$, $p<.0001$, suggesting larger effects were calculated using dependent t-tests, d=.40, than using means and standard deviations, d=.19. The random-effect model revealed no significant difference, $Q(1,k=81) = 0.66$, ns.

Type of two-group design. Next, within the two-group designs we compared samples from evaluations that employed random assignment of students to conditions with designs that used nonequivalent or intact groups. Both the fixed-effect homogeneity analysis, $Q(1,k=44) = 1.78$, ns, and the random-effect homogeneity analysis revealed nonsignificant difference, $Q(1,k=44) = 0.88$, ns. However, *the average d-index for samples that employed random assignment was d=.14, and this effect was significantly different from zero using either model of error.* For nonequivalent-group designs the average effect was d=.05 under fixed-effect assumptions and d=.08 under random-effect assumptions and neither model led to a rejection of the null hypothesis.

Because studies that employ random assignment are considered most trustworthy, we next examined in more detail the five evaluations that used this design feature. Geis (1968) conducted the first evaluation of a summer program that randomly assigned students to treatment and control conditions. A summer program in reading was provided for students entering 1st

Curry (1990) reported an evaluation of a seven-week, residential program conducted at Nichols State University. The program provided remedial education, job training, and counseling and was designed to prevent dropping out among 30 at-risk 8th graders identified by their teachers as at-risk. Four measures, reading and math achievement, self-esteem, and locus of control, were taken. During the program, participants made gains in all areas while control-group students showed declines. A six-month follow-up revealed the treated students maintained the gains they evidenced at the end of the program. The overall effect of the treatment was d=.29.

Matching procedures. The results comparing one-group and two-group designs and comparing random assignment and nonequivalent-group designs suggested that interventions using nonequivalent groups had the smallest estimates of the effect of summer school. One explanation for this finding might be that these studies chose control students who were poorly matched with summer school participants. More precisely, the intact control groups might have contained samples of students who were achieving at higher levels relative to treated students prior to summer school and this difference might have suppressed the estimate of the positive effect of summer school.

To test this possibility, we next examined whether, among nonequivalent-group designs, samples that were chosen using some attempt to match students revealed different effects from samples that were not matched. The homogeneity analyses revealed that the use of matching procedures was not associated with a significant amount of variance in effects whether a fixed-effect, $Q(1, k=33) = 0.00$, ns, or random-effects model was used, $Q(1, k=33) = 0.00$, ns. However, within studies that used matching, the average d-index in samples that employed academic achievement as the matching variable revealed larger fixed-effect d-values, d=.11, than did samples that used any other matching procedures, d=-.03, $Q(1, k=24) = 3.94$, $p<.05$. This effect was nearly significant under random-effect assumptions, $Q(1, k=24) = 3.03$, $p<.09$.

Evaluation agent. Next, we examined whether evaluation reports prepared by the same organization that implemented the summer program contained different results from reports prepared by outside agents. Using a fixed-effect model, we found reports that were prepared by the school or school district also responsible for delivering the program, reported larger effects, d=.30, than evaluations prepared by external agents, d=.16, $Q(1,k=99)=38.11$. Using a random-effect model, the comparison was not significant, $Q(1, k=99) = 0.01$, ns, probably because of a large increase in the random-effect estimate of the d-index for external evaluations, d=.25..

Size of the evaluation. The next methodological moderator we looked at concerned the size of the evaluation. Under both fixed and random effect assumptions, we found no significant relationship between the number of students who provided data for the evaluation and the magnitude of the summer school effect, for fixed and random models $Q(1,k=109) = 2.27$, ns.

Indicators of treatment fidelity. Another homogeneity analysis revealed significant fixed-effect difference between samples in evaluation reports that stated the implementation of the summer school program was monitored, d=.34, and reports that made no mention of including such a procedure, d=.20, $Q(1,k=99)=37.96$, $p<.0001$. The random-effects model showed no

We examined the mean d-indexes related to grade-levels in three groups roughly encompassing early primary grades, late primary grades, and secondary grades. We noticed evidence of a curvilinear relationship, suggesting summer programs were most effective for the youngest and oldest students. Table 5 reveals that the largest effect sizes were associated with summer programs that served more advanced students, with d-values from both fixed-effect and random-effect models ranging from .29 to .36. In all cases, students in middle grades revealed the smallest effect sizes, with d-values ranging from .14 to .21. Students in the lowest grades fell between the two extremes, with d-values ranging from .19 to .32. Because of this pattern, we conducted homogeneity analyses to determine the significance of any curvilinear relation. We found significant curvilinear relations for all fixed-effect tests, for lowest grade unadjusted d-values, $Q(1,k=96)=31.92$, $p<.001$, for highest grade unadjusted d-value, $Q(1,k=96)=9.78$,$p<.005$, for lowest grade adjusted d-values $Q(1,k=96)=102.84$, $p<.0001$, for highest grade adjusted d-values, $Q(1,k=96)=18.12$,$p<.0001$. One significant and one near significant curvilinear effect was found in the analyses using random-effect assumptions, for lowest grade unadjusted d-values, $Q(1,k=96)=2.77$, $p<.10$, for highest grade unadjusted d-value, $Q(1,k=96)=0.19$,ns, for lowest grade adjusted d-values $Q(1,k=96)=4.05$, $p<.05$, for highest grade adjusted d-values, $Q(1,k=96)=0.31$,ns.

Student sex, The sex of the student proved to be a nonsignificant moderator of d-values, regardless of whether fixed or random models were used with unadjusted or adjusted d-values for fixed-effect unadjusted d-values $Q(1,k=19)=0.16$, ns, for random-effect unadjusted d-values $Q(1,k=19)=0.16$, ns, for fixed-effect adjusted d-values $Q(1,k=19)=2.81$, ns, for random-effect adjusted d-values $Q(1,k=19)=0.22$, ns. Six studies that examined the effect of summer school on males and females separately contributed 16 independent samples to this analysis (with the remaining three samples coming from studies that contained only males). Because of their within-study comparisons, these evaluations deserve a closer examination.

Wasik and Sibley (1969) evaluated a summer program for culturally-disadvantaged students entering 1st grade. The program used token reinforcement and isolation techniques to teach a wide array of carefully-delineated academic skills. Pre-program to post-program changes in standardized intelligence test scores showed a dramatic effect and the effect was larger for males, d=.66, than for females, d=48.

Williams (1977) also conducted a study that looked at the student's sex as a moderator of a program's effect. This evaluation assessed the impact of a summer program meant to improve the reading ability of 4th grade students who were in a remedial reading class. A nonequivalent-group comparison was used. In general, no significant differences were found between students who did and did not attend the program, although there was some evidence that females in the summer school program, d=.30, outperformed males in the program, d=.11.

Abram and Cobb (1979) and Abram and Maurelli (1980), examined two successive years of the effects of a summer program in Owensboro, Kentucky. Analyses considering the sex of the student revealed results that did not replicate from one summer to the next. In 1979, the program had a slightly larger effect on females, d=.19 than on males, d=.10, but in 1980 two of three comparisons showed larger effects for males.

Baxley and Hinton (1971) reported an evaluation of a summer program that targeted language instruction for children of migrant workers attending school in Eloy, Arizona. The program included preservice and inservice training for teachers. In the program, children received meals and engaged in recreational activities as well as language instruction in an experiential setting. Data on only eight students, a random sample of 15% of students in the program, revealed a d-value of 1.64, which was adjusted to 1.50 for use in this analysis.

In sum then, there is robust evidence from the meta-analysis that summer programs have larger positive effects on middle class than low-income students. However, the effect of summer school was positive and greater than zero for both economic groups. Further, evaluations of programs with children of migrant workers, though few in number, also suggest large effects of summer programs.

Achievement label. The final student characteristic we examined as a moderator of the effect of summer school was the achievement label given to the participating students. While all student groups showed significant gains associated with attending summer school, fixed-effects models revealed significant differences between students given different achievement labels. Curiously however, with unadjusted d-values the means suggested that students labeled as learning-disabled or physically or mental impaired appeared to benefit more from summer school, d=.34, than students labeled underachieving, d=.27, failing, d=.23, or at-risk, d=.19, $Q(3,k=99) = 22.01$, $p<.0001$. However, after adjustment for methodological confounds, students labeled as underachieving, d=.36, revealed the largest effect size while d-values for students labeled with learning disabilities, d=.23, or as failing, d=.21, or at-risk, d=.19, varied little amongst themselves, $Q(3,k=99) = 49.72, p<.0001$. While nonsignificant, the analyses using random-effect assumptions were consistent with this latter finding, for unadjusted d-values $Q(31,k=99)=0.62$, ns, for adjusted d-values, $Q(3,k=99)=3.92$,ns.

Conventional wisdom suggests that students with disabilities should benefit considerably from summer school because it may prevent dramatic summer losses. This supposition is reflected in federal law that mandates summer services for many of students eligible for special education. Because of this, we examined more closely evaluations of programs for children with special needs.

The 18 independent samples of students with learning disabilities were drawn from six separate summer programs. Fonzi (1984) examined a summer program for 3rd and 5th grade students with learning disabilities. Eight samples varied in size from 98 to 510. Some students attended the summer program part-time and some full-time depending on their individual educational prescription. Instruction was given in basic skills in a resource room for part of the day and the remainder of the day was spent in a regular classroom. The evaluation used the statewide testing battery in math and communications (reading and writing) as the outcome measure. Students who attended summer school performed better than their counterparts who did not attend but the effect was not consistent across all grade levels. The d-values ranged from d=-.24 to .67. Also, Fonzi broke out the results of the summer school program according to reading and mathematics stanines. This permitted us to examine the effects of summer school for students with different levels of learning deficiency. Students in the four samples who scored in the lowest stanines (1-3) revealed summer school effects of d=-.13, -.07, .18, and .18. Students in

ns. The number of years the program had been in existence revealed one significant effect. When fixed-effect assumptions were used and d-values were adjusted to remove correlation with methodological factors, programs evaluated in their first year, d=.29, had larger effects than programs evaluated in later years, d=.23, $Q(1,k=99)=8.31, p<.005$. No other analysis of the year of the program came close to producing a significant effect, for fixed-effect unadjusted d-values $Q(1,k=99)=0.15$, ns, for random-effect unadjusted d-values $Q(1,k=99)=0.15$, ns, for random-effect adjusted d-values $Q(1,k=99)=0.23$, ns.

Type of community. The type of community in which the program operated was a strong predictor of summer school outcomes, revealing significant effects under both fixed and random assumptions with unadjusted d-values and near-significant effects with adjusted d-values, for fixed-effect unadjusted d-values $Q(1,k=73)=11.27$, $p<.001$, for random-effect unadjusted d-values $Q(1,k=73)=4.97$, $p<.03$, for fixed-effect adjusted d-values $Q(1,k=73)=3.26$, $p<.08$, for random-effect adjusted d-values $Q(1,k=73)=3.26$, $p<.08$. Summer programs conducted in rural communities, suburbs, or small cities revealed substantially larger d-values, with estimates ranging from .34 to .44, than programs carried out in large cities, ranging from .24 to .29, although these programs also had significant positive effects.

Number of schools and classes. Two indicators of the size of summer programs, the number of schools and classrooms encompassed by the program, also suggested a relation with program outcomes, but here the effects were largely confined to analyses using the fixed-effect models. Programs that generally used fewer schools revealed larger effects than programs occupying more schools, for fixed-effect unadjusted d-values $Q(1,k=50)=54.94$, $p<.0001$, for random-effect unadjusted d-values $Q(1,k=50)=5.75$, $p<.02$, for fixed-effect adjusted d-values $Q(1,k=50)=8.27$, $p<.005$, for random-effect adjusted d-values $Q(1,k=50)=1.22$, ns. A median split showed that summer programs conducted in less than 8 schools had larger d-values, ranging from .37 to .48, than programs conducted in 8 or more schools, ranging from .21 to .26. Programs that used fewer classrooms also revealed larger effects under fixed-effect assumptions than programs employing a greater number of classrooms, for fixed-effect unadjusted d-values $Q(1,k=21)=40.21$, $p<.0001$, for random-effect unadjusted d-values $Q(1,k=21)=2.13$, ns, for fixed-effect adjusted d-values $Q(1,k=21)=13.37$, $p<.0005$, for random-effect adjusted d-values $Q(1,k=21)=1.34$, ns. A median split showed that summer programs conducted with less than 8 classrooms had larger d-values, ranging from .42 to .90, than programs conducted with 8 or more classrooms, ranging from .13 to .18.

Both of the moderators related to the size of a program suggested that larger effects were associated with smaller summer programs. Earlier, we pointed out that another indicator of program size, the actual number of students who were in the program, was reported too infrequently to test as a moderator of program outcomes. Only 11 of the 41 reports included information on the size of the entire summer program and most of the remaining programs could not confidently be assumed to be large or small. Of the 11 reports, 5 reports said the program served less than 200 students and 4 said it served more than 1000 students. It might be instructive to examine these latter four evaluations.

effect unadjusted d-values $Q(1,k=99)=18.16,p<.0001$, for fixed-effect adjusted d-values $Q(1,k=99)=61.41$, $p<.0001$, for random effect adjusted d-values $Q(1,k=99)=10.32,p<.005$.

Our examination of outliers highlighted a program that employed individualized instruction (Cramer & Dorsey, 1969). We found twelve other evaluations that stated instruction was given to students individually. Four of these programs involved students with learning or other disabilities (Leviton, 1973, Menousek, 1983, Franklin, 1987; Miller, 1987) and three others have been described because of other design features (Cobb, 1990; Curry, 1990; Seigelman, 1975). The remaining five programs also deserve brief descriptions.

Amorose (1987) reported on a summer program conducted for Columbus, Ohio, public schools that focused on reading and math skills of at-risk high school students. Interestingly, participants were paid to attend classes. The program met for 39 days and both homogeneous and heterogeneous class groupings were used along with individualized instruction. Pretest-posttest scores revealed a d-value of .50. Tam (1987) also evaluated a program in Columbus, Ohio, that ran during 1987. This program was for 2nd graders who were experiencing difficulties in reading. The 24-day program involved 30 minutes of one-on-one reading instruction each day. The resulting d-value was .74.

Brown (1991) evaluated a residential camping program in South Carolina for predominantly Africa-American boys and girls. Camp lasted for three to four weeks and mornings at camp were spent in academic activities. Pretest-to-posttest effects ranged from d=.31 to .48.

Merkel-Keller (1987) described a program for 10th graders in New Jersey. This program also included both academic instruction and paid employment. Three hours of instruction lasted for six weeks and was provided in groups of students as small as five, with intensive teacher input, involvement, and interaction. The resulting d-value was .16.

Finally, Woloshin (1975) examined a New York City summer program meant to overcome reading and math deficiencies in 9th and 10th graders. Three hours of instruction occurred for six weeks. Individualized instruction was preceded by the use of standardized tests to diagnose each student's strengths and weaknesses and instruction was based on these results. Effect sizes were d=.42 and .47. In sum then, both the class size and individualized instruction findings strongly suggest that greater opportunities for attention to individual students is associated with larger positive effects of summer school.

Residential programs. Five evaluations containing eight independent samples described programs that involved students taking up residence away from home during the summer program. The evaluations by Brown (1991), just described because of its use of individualized instruction, and by Curry (1990), described because it used random assignment, revealed positive summer school effects ranging from d=.29 to .48.

A summer program with the goal of preventing rural, economically-disadvantaged youths from dropping out of school was evaluated by Wells, Springer & McCready (1987). The program included academic, vocational, and counseling components. Students remained in

Q(1,k=67)=2.02,ns, for fixed-effect adjusted d-values Q(1,k=67)=0.00, ns, for random-effect adjusted d-values Q(1,k=67)=0.00,ns. However, an examination of a three-way breakout of fixed-effect, unadjusted d-values suggested a curvilinear component, with summer programs lasting less than 60 hours showing an effects of d=.20, programs lasting between 60 and 120 hours a d=.37, or more than 120 hours a d=.21.

Curriculum content. The content of the summer curriculum was significantly or near-significantly related to the summer program outcomes in three of the four analyses, for fixed-effect unadjusted d-values Q(2,k=52)=10.30, p<.01, for random-effect unadjusted d-values Q(2,k=52)=5.97,p<.06, for fixed-effect adjusted d-values Q(2,k=52)=25.72, p<.0001, for random-effect adjusted d-values Q(2,k=52)=3.12,ns. However, the pattern of relationships between effect sizes for programs primarily focusing on reading curricular, on both reading and math curricular, and on multiple subjects does not remain stable across model assumptions and statistical adjustments. While d-values for programs with multiple-subject curricula tend to remain stable, ranging from .19 to .24, the reading and math program outcomes vary in magnitude from d=.25 to .40, and reading-only programs vary most of all, from d=.22 to .43.

An examination of the six studies that contributed the nine samples in the reading curriculum category revealed that seven of the samples had d-values that ranged from .06 to .48, but the remaining two samples provided exceptionally large d-values (from Baxley, 1971, d=1.45 and from Welch, 1990, d=1.50). Thus, the highly variable average estimates are likely a function of the shifting weight given these two studies in the calculation of each estimate.

Tests for Measurement Moderators of Evaluation Outcomes (see Table 8)

Measurement delay. A relatively robust association appeared suggesting that longer delays in measuring a program's effect were associated with smaller d-values, especially under fixed-effect assumptions, for fixed-effect unadjusted d-values Q(1,k=81)=33.64, p<.0001, for random-effect unadjusted d-values Q(1,k=81)=10.40,p<.005, for fixed-effect adjusted d-values Q(1,k=81)=4.83, p<.03, for random-effect adjusted d-values Q(1,k=81)=1.57,ns.

When we examined the five studies that measured the effects of summer school with long delays we found two evaluations, David (1974) and Leviton (1973), included multiple follow-up measures. David (1974) found a sharp decrease in effect sizes from a testing immediately following the summer program, d=.58, to the delayed measurement, d=.18, whereas Leviton (1973) found an increase from a September-October testing, d=.27, to a longer delay, d=.44.

Ward (1989) assessed the long-term effectiveness of North Carolina's Basic Education Summer School Program. The program was instituted as a remediation effort for academically at-risk 3rd and 6th graders. All eligible students from around the state were invited to participate and program participants were matched with nonparticipants of similar school, sex, ethnicity, and pre-program achievement test score. In addition to reading and math, participants were tested for the program's effects on general cognitive abilities and strategic processing of information. Results indicated that the summer program effect was not sustained over two years. However, the researcher cautioned that the evaluation was complicated by missing data, grade retention, and

program revealed an average positive effect on attitudes toward school and teachers, and self-esteem. No academic outcome measures were reported.

Leviton (1973), described above, included measures of student self-esteem and found mixed results, with effects ranging from d=-.04 to d=.32. Ward (1989), just mentioned for inclusion of general cognitive measures, included a measure of student absences during the following school year. The post-program comparisons generally showed participants had more absences than nonparticipants. Brown (1991), Curry, (1990), and Rawson (1992) also contributed effect sizes to the calculation of effects on nonacademic outcomes.

A demonstration of meta-analysis using comparisons generated within studies. The outcome content data set provided us with an opportunity to demonstrate a very powerful meta-analytic technique. Recall that for some moderator variables a single evaluation could contribute more than one effect size if it collected data on more than one moderator category. For example, in our examination of sex differences in summer school outcomes, we were able to identify a few studies that included both males and females and gave separate results for each sex. Similarly, two evaluations were found that tested summer school outcomes with different lengths of delay. These evaluations can be highly informative because they control internally for many of the methodological confounds (known and unknown) that weaken the inferences achievable through meta-analysis. That is, when an evaluation breaks out the effect of the program for males and females separately, we can be relatively confident that many other aspects of the program did not covary with the student's sex. The length of the program, instructional practices, and testing conditions ought to be about the same for both males and females.

This within-study control is even stronger when we consider evaluations that include both a measure of reading and a measure of math achievement. Here, not only are all program-related variations controlled by the within-study multiple measurements but student variations are controlled as well, because each student takes both subject area tests. Therefore, as a final means for examining the outcome content findings we identified each of the evaluations that included both a reading and a math outcome measured on the same students. By looking at these studies separately, we could determine whether the outcome content difference emerged within evaluations in which other components of the program were held constant. There were 21 evaluations including 56 independent samples that used both reading and math outcome measures. All but three of these evaluations have been described already.

A summer program conducted in Milwaukee, Wisconsin (Gousha, 1968), prevailed on teachers to rely heavily on the use of city and neighborhood resources for field trips. Positive effects were found on both math, d=.29, and reading skills, d=.22. Teachers reported a lack of time for instruction and delays or waiting for supplies were the greatest weaknesses of the program.

An Austin, Texas, summer program for 7th- and 8th-grade retainees was evaluated during the summer of 1980 (Porterfield & Eglsaer, 1980). The program's objectives were to improve basic academic skills and decision-making skills. Results revealed that 7th graders made significant gains on math computation skills while 8th graders showed significant gains on three

Second, an evaluation by D'Agnostino and Hiestand (1995) examined a summer program for disadvantaged 4th graders in Chicago, Illinois. This study was unique because it included an observation and categorization of summer school classrooms based on how much focus was placed on advanced or authentic instruction. These types of instruction were defined as including encouragement of student problem-solving, and critical and creative thinking, among other instructional strategies. This evaluation also employed Rasch measurement procedures and hierarchical linear modeling of the program's effect. The evaluators reported that classrooms with a greater focus on advanced learning skills contained students who made greater yearly math and reading gains. Regardless of the instructional distinction, the studies also revealed that summer school students made positive pretest-to-posttest gains in reading and math but did more poorly than a matched control group.

local high school. Put differently, Klein speculated that control group students eventually received benefits similar to those received by students attending the Yale program.

Kulieke (1986) and Olszewski, Kulieke, and Willis (1987) evaluated two separate years of three acceleration programs conducted in the Chicago, Illinios, area. Aimed at gifted middle school students, the programs focused on science and math instruction. They also gave students the opportunity to interact with practicing scientists and to acquire skills associated with application of the scientific method. One program was run at the Argonne National Laboratory, and two at Northwestern University. For the first evaluation, pre-program to post-program changes in attitudes toward science, self-esteem, and the College Board Achievement Test in Biology served as the outcome measures. The effects on achievement scores were very positive but mixed results were found on the nonacademic measures. Similar mixed findings were found during the second evaluation on a measure of global self-worth.

Kolloff and Moore (1989) presented data on the self-concepts of 508 gifted 5th through 10th grade students who took part in three residential summer programs. One program ran at the University of Northern Colorado and two ran at Purdue University. In all three programs, students chose courses from numerous alternatives taught by college faculty. Two measures of self concept revealed d-values of pre-program to post-program change ranging from .22 to .37. No differences in the effect of the programs were found based on the sex or grade level of the student.

Meta-Analysis of Acceleration Program Evaluations (see Figure 1)

As with the meta-analysis of remedial programs, first we Winsorized extreme d-values so that no comparison contributed an effect size larger than d=1.50 (necessary in 4 cases). No sample size was larger than 500 or condition size larger than 250. A total of approximately 2200 students took part in the evaluations.

Of the 60 comparisons, 18 produced negative values. Of 34 independent samples, 9 produced average negative effects across all outcome measures. The average unweighted d-value cross the 34 independent samples was d=.23. The median effect size was d=.25. When the contribution of samples was weighted by their sample sizes the average d-value under fixed-effect assumptions was d=.16. The 95% confidence interval for the weighted d-value ranged from .10 to .21. Under random-effect assumptions, the weighted average d-index was .18 with a 95% confidence interval ranging from .07 to 30. Thus, *we can reject the null hypothesis that summer programs with the goal of accelerating student achievement had no effect. Students in these programs outperformed their own pretest scores or the post-test comparison group by about one-sixth of a standard deviation.*

The fixed-effect homogeneity analysis revealed much greater variation in d-values than would be expected from sampling error alone, $Q(1,k=34)=98.56, p<.0001$. However, our examination of the codes for moderator variables revealed too much missing data and/or widely disparate numbers of independent samples within code categories for us to place great faith in the outcomes of any formal tests of the moderating effects of methodological, student, program, or outcome characteristics. Because they were of high interest, we did run analyses looking at effect

Chapter 7

Narrative and Meta-Analytic Synthesis of Programs with Other and Multiple Goals

Seven evaluations including 20 independent samples and 32 comparisons examined summer programs that had goals other than remediation or acceleration or had multiple goals that were not tested independently. Approximately 3100 students were included in these samples. Table 10 provides some pertinent information on each study.

Narrative Description of Other and Multiple Goal Program Evaluations

The earliest report of such a program we were able to obtain was an assessment of a 1965 Human Development Project carried out in Richmond, Virginia, schools and cosponsored by the Ford Foundation (Tyler, 1966). The program included a variety of activities in the area of language arts, reading, speech, and arts and crafts that were either remedial or enrichment in focus. This evaluation included teacher responses to questionnaires, as well as pre-program and post-program comparisons of achievement test scores, and delayed achievement test comparisons between students at schools who did and did not have the summer program. Generally, teachers rated remedial activities as having the greatest value. Pre-program versus post-program analyses of achievement data revealed significant positive effects on reading and language skills and nonsignificant positive effects on arithmetic. This was not surprising given that math-related activities were not a large component of the summer program. The effect sizes for the two independent samples included in this evaluation were d=.46 and d=.51.Teachers reported they felt student's academic skills improved somewhat more than expected but felt more positive gains occurred in student attitudes than achievement. No differences in pre-program versus post-program scores were found based on the sex of the student but students with IQ scores above 85 were found to benefit more than students with IQ scores below 85. Comparisons between schools that did and did not take part in the project were difficult to interpret because scores were compared for the schools as a whole and it is not clear from the report how comparison schools were chosen.

Perhaps the most thorough and informative evaluation of summer school was provided by Heyns (1978). She examined the impact on 6th graders of a large, voluntary, free-tuition, summer program conducted in 23 different Atlanta, Georgia, public schools in 1972. The program was disproportionately likely to be used by black families, even when the SES of families was controlled. A wide variety of courses was offered on diverse topics, including everything from arts and sports to remedial classes. However, on average, participants' achievement levels were equal or above those of non-attendees and only a small fraction of students were enrolled in order to make up academic work. Attendance varied greatly from school to school, ranging from 55% to 89%. Heyns examined questionnaires completed by students, teachers, principals, and parents along with student scores on a word recognition test, said to be indicative of scores on other tests as well. All parties agreed the program was a success. Teachers invariably mentioned small class sizes, allowing informal and individualized instruction, and greater personal autonomy as the causes of success. Children reported that summer school was fun. Parents praised the program for keeping children off the streets. This was especially true of comments by working mothers,

Finally, Thompson (1989) reported on a Chicago, Illinois, summer program for disadvantaged girls that was undertaken with multiple goals in mind. The program started out as a tutoring, sports, and trade program but then expanded to goals of academic achievement and character development. The character development aspects of the program focused on improving self-image to provide a basis for resisting drug and gang involvement. Students took classes in math and science, communication skills, character development, fine arts, and team sports. The program served primarily African-American and Hispanic students in 7th through 9th grades. A single measure of self-esteem revealed d-values ranging from d=-.57 to 2.89.

Meta-Analysis of Acceleration Program Evaluations (see Figure 1)

The seven evaluations of summer programs with multiple goals contained 32 comparisons taken from 20 independent samples. Five comparisons generated extremely large d-values and were adjusted to d=1.5. However, even after adjustment the average unweighted d-value based on independent samples was .48 but the median was .25.

Examination of the 20 sample effects revealed that only one was negative and 15 fell between d-values of .06 and .65. The fixed-effect model average weighted d-value was .26 with a 95% confidence interval ranging from .22 to .30. The random-effects estimate was d=.38, ranging from .25 to .51. Thus, *we can reject the null hypothesis that summer programs with goals other than remediation and acceleration or with multiple goals had no effect on participant. Students in these programs outperformed their own pretest scores or the post-test comparison group by about one-quarter of a standard deviation.*

The fixed-effect homogeneity analysis revealed much greater variation in d-values than would be expected from sampling error alone, $Q(1,k=20)=153.31, p<.0001$. More so than with the evaluations of programs for acceleration, moderator variables revealed too much missing data and/or widely disparate numbers of independent samples within code categories for formal tests of the moderating effects of methodological, student, program, or outcome characteristics.

students who, on average, were performing better than summer program participants prior to the program's implementation. Therefore, the imperfect-matched control group studies likely contained a selection bias that mitigated the estimate of the summer program effect. One of the smallest average effects appeared in evaluations that compared nonequivalent groups with or without an attempt to ensure their similarity to summer school attendees, d=.05. Within these designs, evaluations with better controls generally showed larger effects of summer school. Specifically, evaluations that matched on achievement showed a more positive effect of summer school, d=.11, than evaluations that matched on other variable, d=-.03.

Given that both pretest-posttest and nonequivalent-group designs are open to bias, it is important to examine carefully the results of evaluations that employed random assignment of students to summer school and control conditions and then measured outcomes for the two groups simultaneously. The evaluations that used random assignment revealed an average effect size, d=.14, that was somewhat smaller than the overall average effect size, although it was still significantly different from d=0 under both fixed-effect and random-effect assumptions. Therefore, in our summary statement about the literature, we chose to set the lower expectation of the summer school effect at the estimate generated by evaluations using random assignment (one-seventh of a standard deviation) and the upper expectation at the value generated by the weighted average of all evaluations (one-quarter of a standard deviation). These values seem quite reasonable, and perhaps a bit conservative, in the context of the entire constellation of estimates we derived.

We should also be clear that the overall impact of summer school should be viewed as an average effect found across a diverse group of programs evaluated with a wide variety of methods. As our moderator analyses revealed, these variations influence the effect size estimate in significant ways, to be address below. Put in practical terms, the overall estimate of effect could guide policy decisions at the broadest level, say by federal or state policymakers. However, a local official about to implement a specific summer program for a particular type of student may find effects quite different from the overall finding. This official might consult our tables to find estimates more specific to the student and program characteristics most descriptive of their local effort. Generally however, both our overall confidence intervals and those associated with specific categories of programs suggest it is unlikely the positive effect of most particular programs will be zero.

Second, *summer school programs focusing on acceleration of learning or on other or multiple goals also have a positive impact on participants roughly equal to programs focusing on remediation.* Our analyses of these programs revealed positive effects significantly different from d=0 using either fixed-effect or random-effect model assumptions. However, because of the smaller number of evaluations we were reluctant to test for the robustness of these findings across methodological, student, program, and outcome variances.

Third, *summer school programs have more positive effects on the achievement of middle class students than on students from disadvantaged backgrounds.* Students from middle class families showed more positive effects of summer school, ranging from d=.44 to .56 depending on model assumptions and effect size adjustment, than students from disadvantaged backgrounds, ranging from d=.20 to .24. The difference between the economic groups was significant whether or not d-values were adjusted for methodological confounds and whether

across all our analytic strategies while significant class-size effects were restricted to fixed-effect analyses. The reports explicitly stating that instruction occurred in small groups or one-on-one were among those that produced the largest effects of summer programs. Further, those evaluations that solicited comments from teachers about the positive aspects of summer school often suggested that small group and individual instruction were among the program's strengths.

We see no reason why the more general educational literature showing a relation between class size and achievement ought not apply to summer programs as well (Glass & Smith, 1979; Mosteller, 1997).

In addition to these principle conclusions, there are five other inferences we think can be drawn from the research synthesis, but with less confidence. We will list these in order of our general sense of their trustworthiness.

First, *summer programs that required some form of parent involvement produce larger effects than programs without this component.* While this finding was robust across all our analyses, we chose to place it among our more tentative findings because only three reports explicitly stated that the program included a parent involvement component and, while large on average, their effect size estimates varied widely. Still, the broader literature on parent involvement suggests this ought to be a positive program component (Epstein, 1992; Cooper, Lindsay & Nye, 1999) as long as parents have the skills to carry out the roles assigned to them and do so in a constructive manner (Casanova, 1996). The types of parent involvement required in the summer programs that we reviewed, including conferences with teachers, observing their child in class, and reading at home, are certainly constructive and should be manageable by most parents.

Second, *remedial summer programs may have a larger effect on math achievement than on reading.* We base this conclusion largely on the results of our meta-analysis comparing reading and math measures taken as outcomes within the same program. We think this is a powerful technique for uncovering differential effects of any treatment program. Still, we have placed it among our more speculative findings because the within-study difference was significant only under fixed-effect assumptions and because we did not find a corresponding reliable difference when we examined subject matter differences in the curricula of the summer programs.

It is possible to interpret this finding in relation to summer learning loss. Specifically, the meta-analysis of summer loss research (Cooper, et.al., 1996) revealed that students' achievement scores in math show more of a drop during summer than reading achievement scores. The reason for the difference in loss was attributed to the likelihood that practice in reading is more embedded in students' everyday environments outside of school than is practice in mathematics. If this is the case, then control students likely received less practice during summer in math than in reading. Thus, the difference in the experiences of control students may explain the difference in summer school effects.

By highlighting the finding that summer school may be more efficacious for math than reading outcomes, we do not want to leave the impression that promoting literacy should not be a goal of summer programs. Summer school has positive effects on reading as well as math. Further, illiteracy is a strong predictor of negative social behavior in both children and adults.

Fourth, *remedial summer school programs have positive effects for students at all grade levels, although the effects may be most pronounced for students in early primary grades and secondary school than in middle grades.* We had no *a priori* reason to expect such a curvilinear relationship. However, it proved significant for fixed-effect analyses using effect sizes both unadjusted and adjusted for methodological confounds. Therefore, we will speculate that underlying the relationship may be the existence of three largely independent approaches to summer instruction associated with different grade levels. We base this speculation on Albuquerque Public Schools' (1985) description of the results of interviews of teachers following a summer program for all students. The interviews revealed elementary school teachers felt summer school gave them the opportunity to be more creative and to individualize instruction, middle school teachers said they emphasized study and organizational skills more than during regular session, and high school teachers, because of the credit structure, adhered most closely to regular session content. If these differences in approaches to summer school hold generally, we might expect the greatest achievement gains in the earliest and latest grades because it is here that teachers place the greatest emphasis on instruction in subject matter. Summer school in the middle years may place more emphasis on the teaching of subject-related study skills that eventually, but not immediately, have an impact on most outcome measures. Alternatively, it may simply be the case that summer school is most effective in the later grades because this is where the most summer loss takes place (Cooper, et.al., 1996).

Fifth, *summer programs that undergo careful scrutiny for treatment fidelity, including monitoring to insure that instruction is being delivered as prescribed, monitoring of attendance, and removal from the evaluation of students with many absences may produce larger effects than unmonitored programs.* All three indicators of treatment fidelity proved significantly associated with average effect sizes using fixed-effect but not random-effect assumptions. This association may be due to the impact that surveillance can have on the rigor with which programs are delivered. Or, it may simply indicate that the extra care that monitoring of programs implies is associated with other evaluation features, for example care in data collection, related to the evaluation's likelihood of uncovering an effect.

Finally, two findings of the meta-analysis that deserve mention suggested inconsistent or a lack of mediation by student and program characteristic. First, *there is inconsistent evidence regarding whether or how the achievement label given to students is associated with the amount of benefit students derive from remedial summer programs.* Our two analyses using fixed-effect assumptions produced significant differences between achievement-label groups but the pattern of mean differences was not the same. Neither random-effect analysis proved significant. As we noted in the introduction, one impetus for summer school is the federally-mandated requirement that extended year services be available to children with disabilities who may suffer from regression of skills over the summer break (Public Law 94-142). Our results show clear and reliable benefits of summer school for these children, but these benefits appear relatively no greater in magnitude than the benefits for other students.

Second, summer school remedial programs that require attendance appear no less effective, if not more effective, than programs that are voluntary. Contrary to what we would have predicted, we found that programs identified as voluntary revealed smaller effects than programs that were nonvoluntary or that gave us no information on student choice. The finding

treatment but within the same subclass (e.g., How does the effect of summer school compare to other educational interventions?), and (c) share the same treatment or predictor but vary in outcome measure (e.g., Does summer school influence some spheres of learning more than others?). Cooper also suggested that effect sizes need to be interpreted in relation to the methodology used to generate them (e.g., the sensitivity of research designs, the trustworthiness of measures). A similar approach to interpretation of effect sizes is given by Lipsey (1990).

Cohen gave us our first broad indicator that the summer school effect might be considered "average" among educational effects. Today, however, enough meta-analyses have been conducted that it is possible to find numerous related contrasting elements in the literature. For example, Walberg (1986) presented a list of 11 meta-analysis in education along with the average effect size obtained in each. The educational interventions and strategies on the list included programmed instruction, advanced organizers, and ability grouping, to name a few. The revealed effect sizes in these meta-analyses ranged from d=-.12 to .78. Comparing the findings of the summer school meta-analysis to these findings also suggests summer school had about an "average" impact on achievement.

A more recent and exhaustive compendium of meta-analyses was presented by Lipsey and Wilson (1993). They compiled the results of 302 meta-analyses across the fields of education, mental health, and organizational psychology. The mean d-value across all these meta-analyses was d=.50. About three-fourths of the 180 education meta-analyses revealed larger effects than did the summer school meta-analysis. Twenty-four meta-analyses of programs and techniques meant for remediation or to meet the needs of special students produced a range of d-vales from -.15 to 1.04. Finally, after comparing the results of psychoeducational meta-analyses with others in the field of medicine, Lipsey and Wilson concluded "what does seem clear, however, is that in assessing meta-analytic estimates of the effects of psychological, educational, and behavioral treatment, we cannot arbitrarily dismiss statistically modest values (even 0.10 or 0.20 SDs) as obviously trivial" (pg. 1199).

We would add to Lipsey and Wilson's caution another of our own. Just as primary studies use different methodologies that influence their outcomes, meta-analyses can use methods that are more or less conservative. There were two important decisions we made in conducting the summer school meta-analysis that would place our techniques among the more conservative when compared to other meta-analyses (but more reliable as well, we think). First, we included both published and unpublished reports in our meta-analysis. Other meta-analyses include only published reports and Lipsey and Wilson found that effect sizes gleaned from published reports were 36% larger than effects taken from unpublished reports. Second, our procedure of Winsorizing the most extremely positive effect sizes is still quite rare in meta-analysis.

Of course, judgments of "size" are not synonymous with judgments of "merit." The relative merit of different treatments involves considerations in addition to the magnitude of effect. Most critical among these would be (a) the cost of particular treatments and (b) the value placed on the changes the treatment is meant to create. Levin and colleagues have begun the task of establishing the relative cost-effectiveness of educational interventions (Levin, 1987; Levin, Glass, & Meister, 1987). They compared the cost-effectiveness of computer-assisted instruction, cross-age tutoring, reductions in class size, and increases in daily instructional time. First, they

during the school year interact with the summer school effect, and can chart changes in effectiveness as a function of changes in program delivery.

In addition, we think it would be profitable to pursue other possible influences on summer school effects that only rarely have been the topic of evaluation or were not robust predictors in our data set but which it may be premature to abandon at this point. Three of these are particularly interesting.

First and foremost, future research should expand the variety of outcome measures used to explore the impact of summer school. Research to date has rarely gone beyond the use of standard measures of academic achievement. Yet, the list of goals associated with summer school suggests many other possible outcomes. Even programs with primarily academic goals might find collateral effects that are of interest. We suggest special attention be paid to measures of student delinquency over summer, attitudes toward school, self-image, and attendance and discipline problems during the following school year, as well as broad indices of parent satisfaction and family interaction.

Of course, it is critical that when evaluators include multiple outcome measures a clear distinction is made between those measures that are isomorphic with the primary goals of the program and those that measure collateral outcomes. Otherwise, program effects may appear smaller than is either desirable or anticipated. Related to our above discussion of the size of summer school effects relative to other interventions, we suspect that the average effect of summer school exhibited in this database suffers not insubstantially from many researchers measuring outcomes only tangentially related to program goals.

Second, the optimum length of summer programs appears to be an important topic for future investigation. The meta-analysis found no clear relationship between how long a summer program lasted and how successful it was. It seems reasonable to suggest that the optimum length of a program will interact with the goal of the program. For example, it may be that programs meant to deter juvenile delinquency will have optimum lengths that are longer than programs meant to teacher specific subject content. The question of optimum length is important because it has clear implications for the expenditure of resources. For example, if a program can be shortened and still reach its goals then the program may be able to reach more students or do so in smaller groups while expending equal resources.

Third, as we found ourselves relating the summer school and summer learning loss literatures we became increasingly intrigued by the possibility of a connection between positive summer school effects and the closeness of the last day of the summer program to the start of the next school year. We were not able to test this relation because few reports included the actual dates of the summer program and also employed a delayed measure of summer school effects. Applying a summer learning loss perspective suggests that programs immediately following the preceding school year ought to lose some of their benefit if students forget material in the weeks of vacation that follow. Programs that "back up" to the start of the new school year may be subject to less loss. The timing of summer school seems to have little impact on program costs, so this may be a very efficient way to enhance program outcomes. However, possible negative effects on staffing and attendance need to be considered. All the ramifications of scheduling are excellent topics for future study.

numerous the matching variables, the more confidence we can place in conclusions. Of these matching variables, motivation is clearly the most difficult to measure, as it often requires a testing done especially for the evaluation effort. Therefore, evaluators might consider a host of proxies for motivation that have rarely been used in previous studies, including prior attendance rate, participation in extra-curricular activities, and discipline records.

While district averages provide the most convenient comparison, we have seen that they often can give misleading results. Still, when certain characteristics of summer programs and participants are known, relative to a school district as a whole, we can make predictions about such comparisons that, if borne out, can enhance confidence in findings. Cohort analysis, in which students are compared to students from past years, can also be included in the same investigation as any of the other matching procedures.

Third, to reiterate a point made above, we would suggest that evaluators begin paying much more attention to cost analyses of summer schools. Per-pupil expenditures associated with the operation of summer programs, along with estimates of program effects, could provide invaluable information for policy makers who must decide between the relative effectiveness of two programs meant to accomplish the same goal. For example, assume a school district has instituted a policy that students will no longer be allowed to move on to the next grade unless they are reading at grade level. The district must then decide whether to provide a summer reading program to below-grade readers or to retain them in their current grade. During the first year of operation, the district finds that 50% of students in the summer program reached the grade level criteria. In order to determine whether the program was cost-effective, the district compares the $2000 cost per student of the summer program to the $6000 cost per student of an additional school year. Under such constraints the program would be deemed cost-effective because the "break even" point would be 33% of summer school students meeting the criteria.

Of course, this is a grossly over-simplified example. However, tight educational budgets and calls for greater accountability concerning program cost-effectiveness dictate that such estimates begin to be a part of the data regularly collected when we evaluate educational programs

Finally, we think there is an important place for qualitative research designs in summer school evaluations. For example, ethnographic studies of how summer school classes operate could shed much light on the grade-level differences uncovered in the meta-analysis. Case studies of children and families participating and not participating in summer programs could reveal many unexpected effects of programs, both in terms of behaviors that summer school fosters and behaviors that it replaces. Focus groups of students, parents, and teachers, could also reveal latent effects as well as aspects of summer program design and delivery that might otherwise go unnoticed by evaluators.

The Implications of Research for Summer School Policies

We would like to conclude by offering some guidelines to policymakers and program implementers concerning the funding, development, and operation of summer schools. The guidelines are in no way inconsistent with the results of our research synthesis and most are based on them. However, on occasion, our proposals go beyond recommendations that have identifiable research underpinnings. When this occurs, we have drawn on inspiration gleaned

that flexible delivery systems may lead to important contextual variations that significantly improve the outcomes of summer programs. Therefore, policymakers ought to resist the temptation to micromanage programs and give local schools and teachers leeway in how to structure and deliver programs.

Finally, *policy makers should require rigorous formative and summative evaluation of program outcomes*. Funds should be set aside for this specific purpose. Exactly how prescriptive policies requiring evaluation should be ought to depend on the funding mechanism (e.g., block versus competitive grants) and how distant the source of funds is from the operation of the program (e.g., federal versus local district mandates). Credible evaluations provide the accountability that is called for to justify expenditure of public funds. Also, while it always sounds self-serving for researchers to call for more funds for research, we hope that our attempt at integrating the research on summer schools has revealed both the benefits of research and the need for more credible research. Policy makers can make a substantial contribution to future data-based decision making by requiring and providing funds for systematic, ongoing program evaluation.

Guidelines for program implementers. There are numerous suggestions for implementation that can be gleaned from the research synthesis. For example, implementers should weigh the trade-off between the length of the program and the number of students served and how this trade-off is influenced by the program's goals. Implementers should consider ending programs close to the start of the new school year. Beyond the quantitative research results, some other suggestions were mentioned multiple times in the literature.

Surveys of teachers that accompany evaluations often point to a lack of planning time and late-arriving program materials as two of the most severe impediments to the success of a summer program. Thus, just as policy makers need to provide stable and continuing sources of funds for summer schools, program implementers need to plan early. We suspect that the pragmatics of program operation will take on a higher priority as summer schools are seen less as "add ons" and more as integral parts of the array of services provided by schools.

Related to planning is the need for program implementers to provide continuity from year to year. Priority for staffing should be based on past participation in the summer program itself so that teachers, administrators, aides, and support staff who took part in past years are given the first opportunity to be involved again. Evaluations should be used to continue successful elements of a program, from site locations to program content, and to discontinue unsuccessful ones.

Finally, program implementers might also consider integrating summer staff development activities for teachers with the teaching of summer school. The relatively small classes and relaxed atmosphere that many summer programs provide could make them an ideal laboratory for teachers to experiment with new curricula or pedagogical approaches. For example, teachers might learn about and discuss a new teaching strategy in the morning and then practice the approach using an afternoon summer school class. The coupling of staff development and summer teaching might also increase the pool of teachers interested in taking part.

The Running Start Summer Program Component. As one example of how to break routine thinking about summer school, we offer a summer school program component we call the

References

Note: After each reference a * indicates those studies included in the direction only analysis and ** indicates those studies included in the meta-analysis.

Abram, M. J. & Cobb, R. A. (1979). Evaluating a summer school program for elementary students achieving below grade level. Bowling Green, KY: Professional Development Center Network. (ERIC Document Reproduction Service No. ED 197 971).**

Abram, M. J. & Maurelli, J. A. (1980). Does attendance in summer school increase reading and math achievement of elementary students? Owensboro, KY: PREPS Research Project. (ERIC Document Reproduction Service No. ED 209 269).**

Adams, M.J. (1991). Beginning to read: Thinking and learning about print. Cambridge, MA: MIT Press.

Agnew, A. T. (1975). The effects of a summer communication skills program upon selected language arts skills and dimensions of the self concept of disadvantaged negro pupils. Unpublished doctoral dissertation, New Mexico State University, Las Cruces.*

Albuquerque Public Schools. (1985). What I did instead of summer vacation: A study of the APS summer school program. Albequerque, NM: Albuquerque Public Schools. (ERIC Document Reproduction Service No. ED 281 932).

American Psychological Association. (1994). Publication manual (4th ed.). Washington, DC: Author.

Amorose, A. M. (1987). Summer academic skills enhancement program, summer 1987. Final evaluation report. Columbus, OH: Columbus Public Schools, Department of Evaluation Services. (ERIC Document Reproduction Service No. ED 296 058).**

Ascher, C. (1988). Summer school, extended school year, and year-round schooling for disadvantaged students. ERIC Clearinghouse on Urban Education Digest, 42, 1-2.

Ashton, B. D. (1983). An evaluation of the process and product of the 1981 let's read summer school. Unpublished doctoral dissertation, The University of Michigan.**

Association of California School Administrators (1988). A primer on year-round education. Sacramento, CA: Author.

Austin, G.R., Rogers, B.G. & Walbesser, H.H. (1972). The effectiveness of summer compensatory education: A review of the research. Review of Educational Research, 42, 171-181.

Barnett, V. & Lewis, T. (1978). Outliners in statistical data. Chicester, England: Wiley.

Barrett, M.J. (1990, November). The case for more school days. The Atlantic Monthly, ??, 78-106.

Baxley, D. M. & Hinton, M. (1971). The eloy story. A report from the eloy elementary school summer migrant program for kindergarten through second grade level children. Phoenix,

Cooper, H., Nye, B., Charlton, K., Lindsay, J. & Greathouse, S. (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. Review or Educational Research, 66, 227-268.

Cramer, W. & Dorsey, S. (1969). A summer developmental reading program for rural students. The Reading Teacher,22, 710-713, 724.**

Culp, C. R. (1973). An analysis of student achievement in arithmetic and reading as a result of summer school instruction in a selected Indiana city. Unpublished doctoral dissertation, Indiana University.**

Curry, B. A. (1990). The impact of the nicholls state-youth opportunities unlimited program as related to academic achievement, self-esteem, and locus of control. Louisiana. (ERIC Document Reproduction Service No. ED 333 321).**

Curry, J. G., Griffith, J., Washington, W., & Williams, H. (1996). Title 1/title 1 migrant evaluation report, 1995-96. Publication Number 95.02. Austin, TX: Austin Independent School District. (ERIC Document Reproduction Service No. ED 405 447).*

D'Agostino, J. & Hiestand, N. (1995, April). Advanced-skill instruction in chapter 1 summer programs and student achievment. Paper presented at the meeting of the American Educational Research Association, San Francisco, CA. (ERIC Document Reproduction Service No. ED 390 929).**

Dailey, C. J. (1979). Colorado migrant education program. 1977-78 evaluation report. CDE evaluation series, No. 80-4. Denver, CO: Colorado State Department of Education. (ERIC Document Reproduction Service No. ED 192 979).*

David, J. L. (1974). Summer study: A two-part investigation of the impact of exposure to schooling on achievement growth. Unpublished doctoral dissertation, Harvard University.**

DeLing, E. (1972). In-depth evaluation of oral language instruction in the 1971 migrant education summer program. Mount Pleasant, MI: Central Michigan University. (ERIC Document Reproduction Service No. ED 064 014).*

Donaldson, W. S. (1989). Summer Academic Skills enhancement program. Final evaluation report. Columbus, OH: Columbus Public Schools, Dept. of Evaluation Services. (ERIC Document Reproduction Service No. ED 319 860).*

Donaldson, W. S. (1990). Summer academic skills enhancement program 1990. Columbus, OH: Department of Evaluation Services. (ERIC Document Reproduction Service No. ED 327 608).*

Donaldson, W. S. (1992). Fast track final evaluation report summer academic skills enhancement program. Columbus, OH: Department of Evaluation Services. (ERIC Document Reproduction Service No. ED 354 286).*

Doss, D., Totusek, P., Curtis, J., Pankratz, C., Walker-Wheatley, K., Washington, W., & Shaw, L. (1979). Washington, D. C.: Department of Health, Education, and Welfare. (ERIC Document Reproduction Service No. ED 188 823).**

Dougherty, J. W. (1981). Summer School: A New Look. Bloomington, In: Phi Delta Kappa.

Hansen, J. B., Yagi, K., & Williams, G. L., (1986). Elementary Summer School, 1985 Program in the Portland Public Schools. Portland Public Schools, OR: Dept. of Research, Evaluation. (ERIC Document Reproduction, Service No. ED 273 662).*

Hayes, E. M. & Kerr, T. H. (1970). An inter-disciplinary evaluation of a summer program for the rural disadvantaged youth in Nelson County, Virginia. Final Report. Washington, D.C.: Office of Education, Bureau of Research. (ERIC Document Reproduction Service No. ED 046 599).**

Hedges, L. B. & Olkin, I. (1985). Statistical methods for meta-analysis. Orlando, FL: Academic Press.

Hedges, L.V. & Vevea, J.L. (1998). Fixed and random effects models in meta-analysis. Psychological Methods, 3, 486-504.

Heyns, B. (1978). Summer Learning and the Effects of Schooling. New York, NY: Academic Press.**

Heyns, B. (1986). Summer programs and compensatory education: The future of an idea. New York. (ERIC Document Reproduction Service No. ED 293 906).

Hyman, J. E. (1988). The effects of a remedial six-weeks summer school program on the achievement in and attitude toward reading and mathematics in grades four through eight. Unpublished doctoral dissertation, University of South Carolina.**

Karnes, F. A., Meriweather, S., & D'Llio, V. (1987). The effectiveness of the leadership studies program. Roeper Review, 9, 238-240.**

Kashmuk, J. & Yagi, K., (1985). Summer school 1984 program in the Portland public schools. 1983-84 evaluation report. Portland, OR: Portland Public Schools. Research and Evaluation Dept. (ERIC Document Reproduction Service No. ED 255 551).*

Klein, E. B. & Gould, L. J. (1970). Evaluation of the Yale summer high school: An experimental demonstration program of compensatory education for disadvantaged high school students. A two year follow up and evaluation. New Haven, CT: Yale University. (ERIC Document Reproduction Service No. ED 047 082).**

Klibanoff, L. S. & Haggart, S. A. (1981). Technical report #8 from the study of the sustaining effects of compensatory education on basic skills. Mountain View, CA: RMC Research Corporation. (ERIC Document Reproduction Service No. ED 213 789).**

Kolloff, P. B. & Moore, A. D. (1989). Effects of summer programs on the self-concepts of gifted children. Journal for the Education of the Gifted, 12, 268-276.**

Kulieke, M. J., (1986). Research design issues in the evaluation of programs for the gifted: A case study. Journal for the Education of the Gifted, 9, 193-207.**

Levin, H. M. (1987). Cost-benefit and cost-effectiveness analysis. New Directions for Program Evaluation, 34, 83-99.

Levin, H. M., Glass, G. V., & Meister, G. R. (1987). Cost-effectiveness of computer-assisted instruction. Evaluation Review, 11, 50-72.

Moss, J. R. & Griffen, B. (1986). Utah Migrant Education Annual Summer Evaluation Report, 1986. Salt Lake city, UT: Utah State Office of Education. (ERIC Document Reproduction Service No. ED 280 669). *

Mosteller, F. (1997). Smaller classes do make a difference in early grades. Harvard Education Newsletter,

National Commission on Excellence in Education. (1983). A nation at risk: The imperative for educational reform. Washingto DC: U.S. Department of Education.

New York City Board of Education (1992). Chapter 1-funded summer programs: Institute for career exploration (ICE), the basic skills programs, the basic reading program, and project YOU (youth opportunities unlimited). OREA report. Brooklyn, NY: Office of Research, Evaluation, and Assessment. (ERIC Document Reproduction Service No. ED 357 128).*

North Carolina State Department of Public Instruction, (1988). Report of student performance. The 1987 basic education program summer school. Raleigh, NC: North Carolina State Dept. of Public Instruction. (ERIC Document Reproduction Service No. ED 295 949).*

Oliphant, L. A. (1988). A two-year analysis of the relationship between summer school participation and non-participation and the reading and mathematics achievement of primary school children. Unpublished doctoral dissertation, The University of San Francisco.**

Olszewski, P., Kulieke, M. J., & Willis, G. B. (1987). Changes in the self-perceptions of gifted students who participate in rigorous academic programs. Journal for the Education of the Gifted, 10, 287-303.**

Opuni, K. A. Beating the odds summer school: A dropout prevention program for at-risk students (1990). Houston, TX: Houston Independent School District. (ERIC Document Reproduction Service No. ED 327 628).**

Ouellette-Howitz, J. & Murray, T. (1998). Annual report to the board of education minneapolis public schools. Minneapolis, MN: Minneapolis Public Schools. Unpublished manuscript.*

Overton, R.C. (1998). A comparison of fixed-effects and mixed (random-effects) models for meta-analysis tests of moderator variable effects. Psychological Methods, 3, 354-379.

Petro, J. R., Pimentel, D. C., Hinkle, B., Burnes, J., & Stewart, D., (1993). On the move: Migrant education 1991-92. Denver, CO: Colorado State Dept. of Education. Planning and Evaluation Unit. (ERIC Document Reproduction Service No. ED 364 383).*

Petro, J., Pimentel, D. C., Hinkle, B., Burnes, J., & Stroup, K., (1994). On the move: Migrant education 1992-93. Denver, CO: Colorado state Department of Education. (ERIC Document Reproduction Service No. ED 383 500).*

Siegelman, M. (1975). Summer mathematics remediation for incoming pupils. 1975 high school umbrella #2. Brooklyn, NY: Board of Education of the City of New York Office of Educational Evaluation. (ERIC Document Reproduction Service No. ED 138 661).**

Sipe, C. L. (1986). Summer training and education program (STEP): The experience of Hispanic participants in the summer of 1985. Philadelphia, PA: Public/Private Ventures. (ERIC Document Reproduction Service No. ED 286 697).*

Sipe, C. L. (1986). Summer training and education program (STEP): Report on the 1986 experience. (ERIC Document Reproduction Service No. ED 283 027).*

Sipe, C. L., Grossman, J. B., & Milliner, J. A. (1988). Summer training and education program (STEP). Report on the 1987 experience. Philadelphia, PA: Public/Private Ventures. (ERIC Document Reproduction Service No. ED 300 479).**

Smith, C. M. (1972). An exploratory study of the effects of compensatory education on the reading and mathematics achievement of intermediate grade pupils. Paper presented at the meeting of the American Educational Research Association, Chicago, IL.

Smith, F. L. (1992). Nature-computer camp 1991. Chapter 2 program evaluation report. Washington, DC: District of Columbia Public Schools. (ERIC Document Reproduction Service No. ED 352 345).*

Steinmiller, G. & Duncan, L. B. (1991). Arkansas youth opportunities unlimited follow-up. Final. Arkadelphia, AR: Henderson State University. (ERIC Document Reproduction Service No. ED 361 555).*

Steinmiller, G. & Steinmiller, R., (1993). Youth opportunities unlimited: The results of a three year study. AR. (ERIC Document Reproduction Service No. ED 358 996).*

Stevenson, H. W. & Lee, S. Contexts of achievement. Monographs of the Society for Research in Child Development, 55 (1).

Strike, K., & Poser, G. (1983). Types of syntheses and their criteria. In S. Ward & L. Reed (Eds.), Knowledge structure and use: Implications of synthesis and interpretation (pp. 343-361). Philadelphia: Temple University Press.

Taggert, J. B., Mecham, S. K., & Ortega, J. (1991). Utah migrant education annual summer evaluation report. Salt Lake City, UT: Utah State Office of Education. (ERIC Document Reproduction Service No. ED 244 726).*

Tam, H. P. (1987). Language development component, CLEAR-reading of evaluation services. Columbus, OH: Columbus Public Schools, Department of Evaluation Services. (ERIC Document Reproduction Service No. ED 290 126).**

Texas Education Agency (1982). Report on ESEA, Title I migrant, 1980 – 81 (Texas). Austin, TX: Texas Education Agency. (ERIC Document Reproduction Service No. ED 218 040).*

Thompson, D. R. (1989). Metro achievement program summer 1988. External evaluation report. (ERIC Document Reproduction Service No. ED 317 651).**

Worsnop, R.L. (1996). Year-round schools: Do they improve academic performance? CQ Researcher, 6, 433-456.

Yinger, J. M., Ikeda, K., & Laycock, F. (1970). Middle start supportive interventions for higher education among students of disadvantaged backgrounds. Oberlin, OH: Oberlin College. (ERIC Document Reproduction Service No. ED 047 659).*

Young, V. M. (1971). Summer school for poor readers – a Title I model project. The Reading Teacher, 24, 526-531, 585.

Contributors

Harris Cooper is Professor of Psychology at the University of Missouri-Columbia. Dr. Cooper's research interests follow two paths. His book Synthesizing Research: A Guide for Literature Reviews is intended to teach advanced undergraduate and graduate students how to conduct all phases of research synthesis. He is also the editor, with Larry Hedges, of the Handbook of Research Synthesis. Dr. Cooper and his students have published reviews in varied journals including the Psychological Bulletin, the Journal of Personality and Social Psychology, the Review of Educational Research, and Developmental Medicine and Child Neurology.

Dr. Cooper is also interested in the application of social psychology to educational policy issues. His research synthesis on homework was published as both a research monograph and as a policy guide for school administrators. He has also published over two dozen journal articles and book chapters in the area of teacher expectation effects.

In 1984, Dr. Cooper received the first Raymond B. Cattell Early Career Award for Programmatic Research from the American Educational Research Association (AERA). He is a Fellow of the American Psychological Association and the American Psychological Society. Dr. Cooper is an Associate Editor of Social Psychology of Education, and an advisory editor for the Psychological Bulletin and the Journal of Experimental Education.

From 1992-1998, Dr. Cooper was an elected member of the Columbia, Missouri, Board of Education, overseeing a school district with 16,000 students and a $100,000,000 annual budget. His article entitled "Speaking Power to Truth: Reflections of an Educational Researcher After Four Years of School Board Service" (Educational Researcher, January, 1996) won the 1997 AERA Award for Interpretive Scholarship. Dr. Cooper is the father of two teenagers and the spouse of a 3rd grade teacher.

Kelly Charlton is a Visiting Assistant Professor in the Psychology Department at the University of Missouri. Her research interests include the social psychology of education, and intra- and intergroup relationships.

Jeffrey C. Valentine is a graduate student in the Program in Social Psychology at the University of Missouri. His research interests include the social psychology of education, primary prevention programs for adolescents, and aggression.

Laura Muhlenbruk is a graduate student in the Program in Social Psychology at the University of Missouri. Her research interests include the social psychology of education.

(Table 1 continued)

Author	Year of report	Location	Sample size	Grade or age	Student characteristics	Type of comparison	Outcome measure	Direction of results
Mangino	1983	Texas	679	7th – 8th	Failing	Post-post	Promotion	2 ap
Pfeifer	1985	Oregon	22	7th – 8th	Failing	Post-post	English, math, Science, geography, health, & world culture	2ap
Kashmuk	1985	Oregon	129	3rd – 7th	Failing	Pre-post	Math, reading, & language	5 ap 2 mp 1 mn 1 an
Burnes	1985	Colorado	2,055	k – 12th	Migrant	Pre-post	English, math, & communication	1 e
Hansen	1986	Oregon	225	3rd – 7th	Failing	Pre-post	Math, reading, & language	3 ap 4 mp 3 mn
Rachel	1987	Louisiana	16,261	2nd – 5th	Below grade	Post-post level	Retention	4 an
Carroll	1987	Connecticut	2,242	k – 8th	Below grade	Pre-post level	Math, reading, & language	3 ap

(Table continues)

(Table 1 continued)

Author	Year of report	Location	Sample size	Grade or age	Student characteristics	Type of comparison	Outcome measure	Direction of results
Rose	1992	Colorado	2,343	2nd – 12th	Migrant	Pre-post	Math & reading	11 ap
Donaldson	1992	Ohio	58	9th – 12th	Below grade	Pre-post level	Math, reading, & language	1 ap
Petro	1993	Colorado	1,984	2nd – 12th	Migrant	Pre-post	Math & reading	2 ap
Petro	1994	Colorado	1,069	2nd – 10th	Migrant	Pre-post	Math & reading	1 ap 1 e
Curry	1996	Texas	172	?	Below reading	Pre-post level	Math & reading	2 an 2 ap
McNeeley	1996	Illinois	175	3rd – 6th	Below reading level	Pre-post	English	1 ap
Ouelette	1998	Minnesota	3,892	4th – 6th	?	Pre-post	Math & reading	5 ap

Note. ap = all positive; mp = mostly positive; e = even; mn = mostly negative; ap = all negative

(Table 2 continued)

Author	Year of report	Location	Goal	Sample size	Grade or age	Student characteristics	Type of comparison	Outcome measure	Direction of results
Robbins	1991	Indiana	Increase positive attitudes	164	1st – 6th	Below grade level	Pre-post	Vocabulary, & comprehension	1 ap 4 e 1 an
Smith	1992	Washington, DC	Enrichment	406	6th	?	Pre-post	General knowledge	1 ap
Steinmiller	1993	Arkansas	Increase positive attitudes	?	9th – 12th	?	Pre-post	GPA	1 ap 1 mp 1 e

Note. ap = all positive; mp = mostly positive; e = even; mn = mostly negative; ap = all negative

Table 3 (continued)

Cramer	1969 (1966)	West Alexander, OH	103	4th – 8th	Below grade level	Pre-post	Reading	1.50 – 1.50(5)
Baxley	1971	Eloy, AZ	8	k – 2nd	Failing	Pre-post	Math & reading	1.45(1)
Culp	1973 (1972)	IN city	226	3rd	Below grade level	Post-post and pre-post	Math & reading	-.22 - -.08(2)
Leviton	1973	Suburban, MN	60	1st – 3rd	Learning disabled	Post-post and pre-post	Attitudes, math, & reading	.10 - .64(3)
David	1974 (1971)	Four cities	137	k-1st	At risk	Post-post and pre-post	Math, reading, & sounds	.12 - .79(4)
Bergeth	1975 (1975)	Minneapolis, MN	123	3rd – 4th	Underachieving	Pre-post	Reading	.03(1)
Siegelman	1975 (1975)	New York, NY	2835	9th – 10th	Below grade level	Pre-post	Math	.36 - .42(2)
Toledo	1975 (1975)	New York, NY	2808	9th	Below grade level	Pre-post	Reading	.06(1)
Woloshin	1975 (1975)	New York, NY	904	9th – 10th	Below grade level	Pre-post	Reading	.43 - .47(2)

(table continues)

Table 3 (continued)

Amorose	1987 (1987)	Columbus, OH	137	8th – 12th	At risk	Pre-post	Math & reading	.50(1)
Franklin	1987	Chicago, IL	55	1st – 7th	Other disability	Post-post	Math & reading	.15(1)
Merkel-Keller	1987 (1985)	Five Urban Districts in NJ	80	10th	Below grade level	Pre-post	Math, reading, & writing	.16(1)
Miller	1987 (1986)	New York, NY	1625	1st – 6th	Mild Handicap	Pre-post	Math & reading	.37 - .50(4)
Sipe	1988 (1987)	5 large urban cities	5549	9th – 10th	Below grade level	Post-post and pre-post	Math & reading	.01- .10(2)
Tam	1987 (1987)	Columbus, OH	26	2nd	Below grade level	Pre-post	Reading	.74(1)
Wells	1987	Northern LA	80	7th – 10th	At risk	Pre-post	Math & reading	.77(1)
Hyman	1988 (1987)	SC	149	4th – 8th	Failing	Pre-post	Attitudes, math, & reading	.11 - .55(5)
Ward	1989 (1986)	NC	471	3rd & 6th	Failing	Post-post	Absences, CAT, math & reading	-.16 - .03(2)

(table continues)

Table 4
Methodological Moderators of Remedial Summer School Evaluation Outcomes

Moderator	k	Fixed Q	Fixed d	Fixed +/-ci	Random Q	Random d	Random +/-ci
[redacted]							
Type of Control	125	35.93			8.74		
pretest-posttest	81		.24	.02		.30	.06
posttest-posttest	44		.07	.05		.09	.07
How Was Pre-Post d Derived?	81	84.91			0.66		
means & sds	65		.19	.02		.27	.07
dependent t	16		.40	.04		.40	.08
Type of Two-Group Design	44	1.78			0.88		
random assignment	11		.14	.11		.14	.11
nonequivalent group	33		.05	.06		.08	.09
Was NE Control Matched?	33	0.00			0.00		
no	9		.06	.12		.07	.21
yes	24		.05	.07		.08	.09
What Was Matched Variable?	24	3.94			3.03		
academic achievement	17		.11	.09		.14	.13
all others	7		-.03	.11		-.03	.11
Who Did Evaluation?	99	38.11			0.01		
externals	55		.16	.04		.25	.08
internals	44		.30	.03		.28	.07
# of Students in Analyses	109	2.27			2.27		
< 90	55		.27	.05		.38	.11
> 90	54		.24	.03		.19	.06
Was Implementation Monitored?	99	37.96			1.03		
no*	73		.20	.03		.23	.06
yes	26		.34	.04		.33	.09
Was Attendance Monitored?	99	14.52			0.79		
no*	63		.22	.03		.24	.06
yes	36		.29	.03		.29	.09
Were Nonattendees Dropped?	99	8.67			0.37		
no*	72		.24	.02		.27	.06
yes	27		.31	.04		.23	.10

Notes. * also includes samples from reports that did not mention this moderator.

Table 6
Program Context Moderators of Remedial Summer School Evaluation Outcomes

		Unadjusted d-indexes						Adjusted d-indexes					
		Fixed			Random			Fixed			Random		
Moderator	k	Q	d	+/-ci	Q	d	+/-ci	Q	d	+/-ci	Q	d	+/-ci
Year of Evaluation	99	0.15			0.15			3.35			1.90		
prior to 1976	36		.28	.03		.33	.10		.33	.03		.33	.10
1976-1981	25		.13	.05		.13	.05		.13	.05		.13	.05
1982 or after	38		.30	.04		.28	.08		.26	.03		.28	.07
Years in Existence	99	0.15			0.15			8.31			0.23		
one year*	69		.26	.03		.31	.07		.29	.03		.27	.07
> one year	30		.25	.03		.19	.07		.23	.03		.24	.05
Type of Community	73	11.27			4.97			3.26			3.26		
other	34		.38	.05		.44	.12		.34	.05		.41	.10
large urban	39		.28	.03		.24	.07		.29	.03		.26	.08
Number of Schools	50	54.94			5.75			8.27			1.22		
< 8	26		.37	.05		.48	.14		.37	.05		.45	.16
> 8	24		.26	.03		.21	.09		.26	.03		.24	.07
Number of Classes	21	40.21			2.13			13.37			1.34		
< 8	11		.64	.13		.90	.36		.42	.13		.67	.35
> 8	10		.18	.05		.13	.12		.17	.05		.16	.10

Notes. * also includes samples from reports that did not mention this moderator.

Table 7 (continued)

Program Feature Moderators of Remedial Summer School Evaluation Outcomes

		Unadjusted d-indexes						Adjusted d-indexes					
		Fixed			Random			Fixed			Random		
Moderator	k	Q	d	+/-ci	Q	d	+/-ci	Q	d	+/-ci	Q	d	+/-ci
Parent involvement?	99	9.96			13.06			9.08			11.33		
no*	91		.25	.02		.24	.05		.26	.02		.24	.05
yes	8		.54	.18		.90	.61		.53	.17		.78	.53
Amount of Instruction	67	7.12			2.02			0.00			0.00		
< 60 hrs 17		.20	.06		.27	.14		.34	.06		.28	.11	
60 to 120 hrs	34		.37	.04		.38	.10		.28	.04		.33	.11
> 120 hrs	16		.21	.05		.20	.11		.29	.05		.34	.09
Curriculum Content	52	10.30			5.97			25.72			3.12		
reading	9		.22	.07		.41	.23		.43	.07		.43	.11
math and reading	16		.40	.04		.37	.09		.30	.04		.25	.13
multiple subjects	27		.20	.04		.19	.06		.22	.04		.24	.06

Notes. * also includes samples from reports that did not mention this moderator.

Table 9
Mean Effect Size by Independent Sample for Acceleration as the Goal

First author	Year of report (program)	Location	Number of students in analyses	Grade	Achievement Label	Type of comparison	Outcome measure	d-index range
Winston	1963	Plainview, NY	102	4th & 5th	Gifted	Post-post	Language, Needs, Science	.31 - .60(4)
Hayes	1970 (1968)	Nelson County, VA	100	9th – 11th	Gifted	Post-post and Pre-post	Self-esteem Science & self-esteem World culture &self-esteem	-.32 - .40(3)
Brody	1984 (1983)	Baltimore, MD	776	7th	Gifted	Post-post	Math & Verbal	-.74 – 1.82(16)
Klein	1970 (1966)	New Haven, CT	72	10th	Average	Post-post	Math, SAT, and class standing	.00(1)
Kulieke	1986 (1984)	Chicago, IL	125	6th – 9th	Gifted	Pre-post	Attitudes about achievement & science	.18 - .74(3)
Olszewski	1987	Chicago, IL	456	7th – 11th	Gifted	Pre-post	Attitudes about achievement	-.10 - .04(4)
Kolloff	1989 (1985)	Northern CO & West Lafayette, IN	508	5th – 11th	Gifted	Pre-post	Self-esteem	.22 - .37(3)

Note. The number of students in analyses is the sum of sample sizes for each independent sample, as such, it includes n's for both pre-post and post-post analyses.

Figure Captions

Figure 1. Schematic Diagram of the Data Analysis Strategy and Overall Results

Figure 2. Stem and Leaf Display of Remedial Summer School Effect Sizes

Figure 3. Models of Changes in Achievement Over Summer as a Function of Student SES,

Subject Areas, and Summer school Attendance

Stem	Leaf
+2.7	15
+2.3	7
+2.2	
+2.1	
+2.0	7
+1.9	
+1.8	7
+1.7	
+1.6	7
+1.5	
+1.4	5
+.9	8
+.8	7
+.7	1479
+.6	467
+.5	0057
+.4	0122234788
+.3	011446677
+.2	23446899
+.1	000011112244555566778899999
+.0	12233568889
-.0	57788
-.1	333367
-.2	24

Note. Effect sizes are based on independent samples (n = 99).