

# Review of Educational Research

<http://rer.aera.net>

---

## Achievement Effects of the Nongraded Elementary School: A Best Evidence Synthesis

Roberto Gutiérrez and Robert E. Slavin  
*REVIEW OF EDUCATIONAL RESEARCH* 1992 62: 333  
DOI: 10.3102/00346543062004333

The online version of this article can be found at:  
<http://rer.sagepub.com/content/62/4/333>

---

Published on behalf of



American Educational Research Association

and



<http://www.sagepublications.com>

Additional services and information for *Review of Educational Research* can be found at:

**Email Alerts:** <http://rer.aera.net/alerts>

**Subscriptions:** <http://rer.aera.net/subscriptions>

**Reprints:** <http://www.aera.net/reprints>

**Permissions:** <http://www.aera.net/permissions>

**Citations:** <http://rer.sagepub.com/content/62/4/333.refs.html>

>> [Version of Record](#) - Jan 1, 1992

[What is This?](#)

## Achievement Effects of the Nongraded Elementary School: A Best Evidence Synthesis

Roberto Gutiérrez  
Robert E. Slavin  
Johns Hopkins University

*A nongraded elementary program is one in which children are flexibly grouped according to performance level, not age, and proceed through the elementary school at their own rates. Popular in the 1950s, 1960s, and early 1970s, the nongraded plan is returning today. This article reviews research on the achievement effects of nongraded organization. Results indicated consistent positive achievement effects of simple forms of nongrading generally developed early: cross-grade grouping for one subject (median ES = +.46) and cross-grade grouping for many subjects (median ES = +.34). Forms of nongrading making extensive use of individualization were less consistently successful (median ES = +.02). Studies of Individually Guided Education (IGE), which used nongrading and individualization, also produced inconsistent effects (median ES = +.11). The article concludes that nongraded organization can have a positive impact on student achievement if cross-age grouping is used to allow teachers to provide more direct instruction to students but not if it is used as a framework for individualized instruction.*

Greek mythology tells us of the cruel robber, Procrustes (the stretcher). When travelers sought his house for shelter, they were tied to an iron bedstead. If the traveler was shorter than the bed, Procrustes stretched him until he was the same length as the bed. If he was longer, his limbs were chopped off to make him fit. Procrustes shaped both short and tall until they were equally long and equally dead.

[Graded systems of school organization] trap school-age travelers in much the same fashion as Procrustes' bed trapped the unwary. (Goodlad & Anderson, 1959, p. 1)

So begins the book that launched one of the most interesting innovations in the history of education: *The Nongraded Elementary School*, by John Goodlad and Robert Anderson. The nongraded elementary school movement was an important force in North American education in the 1960s and early 1970s, even if its major elements were only implemented in a small proportion of schools. The challenge to the traditional age-graded classroom posed by the nongraded concept is one that still has relevance today. More importantly, the nongraded elementary school itself is reappearing in U.S. schools. Recently, the states of Kentucky and Oregon have promoted a shift to nongraded programs, and many districts and schools elsewhere are moving in this direction (Willis, 1991).

A great deal of research has been done to evaluate various forms of the nongraded elementary school, but there are few comprehensive reviews on this topic. McLoughlin (1967), reviewing studies done up to 1966, concluded that most found no differences between graded and nongraded programs in reading, arithmetic, and

language arts performance. In contrast, Pavan (in press), who limited her review of achievement to studies reported between 1968 and 1990, concluded that most comparisons favored the nongraded plan. However, both of these reviews were quite limited. Both simply counted statistically significant findings favoring graded or nongraded programs, paying little attention to the particular forms of nongrading used, the methodological quality of the studies, or the size of the effects.

The purpose of this article is to describe the nongraded elementary school in its earlier incarnations, to systematically review research on the academic achievement effects of nongraded schooling, and to draw inferences from this research for applications of the nongraded ideal in today's schools.

### What Is a Nongraded Elementary School?

The term *nongraded* (or ungraded) elementary school covers a wide range of school and classroom organizational arrangements. Central to the concept is the elimination of traditional grade level designations. Students are grouped according to their level of academic performance, not their ages. Sometimes this grouping is done for just one subject, sometimes it is done for many subjects, and sometimes students are placed in self-contained multiage classrooms according to their reading performance or general ability. For example, a nongraded reading class might contain six-, seven-, and eight-year-old students, all reading at what would ordinarily be considered the second-grade level. Students are allowed to proceed through the grades at their own rates. Some may take longer than usual to complete the elementary grades, while others may complete elementary school in less time than usual. Because a school has classes at many levels, a child who spurts ahead or falls behind can easily be moved to another class appropriate to his or her level. As a result, no child is ever retained or skipped a whole grade at once.

Frequently, the nongraded program applies only to the primary grades (1–3 or K–3) and is called a *nongraded primary school*. This is the main form that is returning today (Willis, 1991). The idea is that all students will have a certain level of academic performance on entering fourth grade but may have taken more or less time to reach this level.

In their original conception, nongraded elementary schools usually incorporated a curriculum structure called *continuous progress*, in which the skills to be learned in such subjects as reading and mathematics are organized into a hierarchical series of levels covering all the grades involved in the plan (usually 1–3 or 1–6). For example, the reading curriculum ordinarily taught in grades 1–3 might be organized in four levels per grade, for a total of 12 levels for the entire nongraded period leading to grade four. In a continuous-progress model, students pick up each year where they left off the previous year. For example, a low achiever who has only completed Level 5 at the end of his or her second year would start with Level 6 at the beginning of his third year (rather than being retained in second grade, as might occur in a graded school). A high achiever who mastered Level 12 at the end of his or her second year might simply go to fourth grade a year early. Goodlad and Anderson (1963) recommended having such a hierarchical structure only for hierarchical subjects (such as reading and math) and using interest groupings, age groupings, or other criteria for subjects (such as social studies and science) that depend to a lesser degree on prerequisite skills.

Beyond the use of flexible, multiage groupings, actual operationalizations of the nongraded elementary school model have varied enormously. At one end of a

continuum of complexity, nongraded organization is essentially equivalent to the Joplin Plan (Floyd, 1954; Slavin, 1987). This is an arrangement in which students are grouped across grade lines for just one subject, almost always reading. For example, at a common reading period all students might move to a class composed of students at the same performance level in reading drawn from different classes and grade levels; a second grade, first-semester reading class might have first, second, and third graders in it. Students move through a continuous-progress sequence of reading levels that cover the material students are expected to learn in all grades involved in the plan. They move as rapidly as they are able to go, taking as much time as they need to master the material. Groupings are reassessed frequently and changed if student performance warrants it.

The main effect of the use of the Joplin Plan is to reduce the number of reading groups taught by each teacher, often to one (i.e., whole-class instruction), thereby reducing the difficulties inherent in managing multiple groups and reducing the need for students to do follow-up activities independently of the teacher.

The Joplin Plan can be described as a nongraded reading program that still maintains an age-graded organization for other subjects. Studies of the Joplin Plan, which was popular in the 1950s and 1960s, do not make it clear what happened when students reached the end of the elementary grades and were reading at a level quite different from their grade level.

In the 1960s, nongraded programs began to resemble more closely the model described by Goodlad and Anderson (1963), which suggested flexible multiage grouping for most or all academic subjects, with continuous-progress curricula for such subjects as reading and mathematics.

When it was first described and implemented in the 1950s and early 1960s, nongraded organization primarily involved changes in grouping patterns, not instructional methods. Teachers in the earlier implementations still overwhelmingly taught students in groups using traditional methods and curricula. Starting in the late 1960s, however, the nongraded plan often absorbed another innovation becoming popular at that time, individualized instruction. Increasingly, descriptions of nongraded schools began to include the extensive use of learning stations, learning activity packets, and other individualized, student-directed activities. In many cases, these individual activities were also combined with tasks students completed in small groups which primarily worked independently of the teacher. Another typical attribute of these forms of nongrading was team teaching. For example, two to six teachers might occupy a section of the school and take joint responsibility for a large group of students, flexibly grouping and regrouping them throughout the day. As time went on, programs of this kind were increasingly implemented in schools without classroom walls and tended to be called open schools rather than nongraded elementary schools (see Giaconia & Hedges, 1982), and, in an introduction to the 1987 reprinting of their 1963 book, Goodlad and Anderson acknowledge the essential commonality between the two approaches.

A good summary of the goals and elements of a fully realized operationalization of the nongraded ideal is adapted by Goodlad and Anderson (1963/1987, pp. xv-xviii) from the dissertation of Barbara Pavan (1972), who was the principal of one of the earliest model nongraded elementary schools and continues to be an important advocate of this approach. It is presented below.

I. *Goals of Schooling*

1. The ultimate school goal is to develop self-directing autonomous individuals.
2. The school should help develop individual potentialities to the maximum.
3. Each individual is unique and is accorded dignity and respect. Differences in people are valued. Therefore the school should strive to increase the variability of individual differences rather than stress conformity.
4. Development of the child must be considered in all areas: aesthetic, physical, emotional, and social, as well as intellectual.
5. Those involved in the school enterprise are colearners, especially teachers and students.
6. The school atmosphere should allow children to enjoy learning, to experience work as pleasurable and rewarding, and to be content with themselves.

II. *Administrative-Organizational Framework*

A. *Vertical Grouping*

7. Each individual works in varied situations where he or she will have opportunities for maximum progress. There are no procedures for retention or promotion, nor any grade levels.
8. A child's placement may be changed at any time if it is felt to be in the best interests of the child's development considering all five phases of development: aesthetic, physical, intellectual, emotional, and social.

B. *Horizontal Grouping*

9. Grouping and subgrouping patterns are extremely flexible. Learners are grouped and regrouped on the basis of one specific task and are disbanded when that objective is reached.
10. Each child should have opportunities to work with groups of many sizes, including one-person groups, formed for different purposes.
11. The specific task, materials required, and student needs determine the number of students that may be profitably engaged in any given educational experience.
12. Children should have frequent contact with children and adults of varying personalities, backgrounds, abilities, interests, and ages.

III. *Operational Elements*

A. *Teaching Materials—Instructional*

13. A wide variety of textbooks, trade books, supplemental materials, workbooks, and teaching aids must be available and readily accessible in sufficient quantities.
14. Varied materials must be available to cover a wide range of reading abilities.
15. Alternate methods and materials must be available at any time so that the child may use the learning style and materials most suitable to his or her present needs and the task at hand (including skill building, self-teaching, self-testing, and sequenced materials).
16. A child is not really free to learn something he or she has not been exposed to. The teacher is responsible for providing a broad range of experiences and materials that will stimulate many interests in the educational environment.

**B. Curriculum (Knowledge)**

17. The unique needs, interests, abilities, and learning rates, styles, and patterns of each child will determine his or her individual curriculum. Conformity and rigidity are not demanded.
18. The curriculum should be organized to develop the understanding of concepts and methods of inquiry more than specific content learning.
19. Process goals will be stressed: the development of the skills of inquiry, evaluation, interpretation, application—the skills of learning to learn.
20. Sequence of learning must be determined by each individual student and his or her teacher, because:
  - (a) no logical or inherent sequence is in the various curriculum areas,
  - (b) no predetermined sequence is appropriate to all learners,
  - (c) individual differences in level of competence and in interest are constantly in flux.
21. Each child will formulate his or her own learning goals with guidance from his or her teachers.

**C. Teaching Methods**

22. Different people learn in different ways.
23. Learning is the result of the student's interaction with the world he or she inhabits. Individuals learn by direct experience and manipulation of their environment: Therefore, the child must be allowed to explore, to experiment, to mess around, to play, and to have the freedom to err.
24. The process is more important than the product. How the child learns is emphasized.
25. All phases of human growth—*aesthetic, physical, intellectual, emotional, and social*—are considered when planning learning experiences for a child.
26. The teacher is a facilitator of learning. He or she aids in the child's development by helping each one to formulate goals, diagnose problem areas, suggest alternative plans of action; by providing resource materials; and by giving encouragement, support, or prodding as needed.
27. Children should work on the level appropriate to present attainment and should move as quickly as their abilities and desires allow them to.
28. Successful completion of challenging experiences promotes greater confidence and motivation to learn than fear of failure.
29. Learning experiences based on the child's expressed interests will motivate the child to continue and complete a task successfully much more frequently than teacher-contrived techniques.

**D. Evaluation and Reporting**

30. Children are evaluated in terms of their past achievements and their own potential, not by comparison to group norms. Expectations differ for different children.
31. Evaluation by teacher and/or child is done for diagnostic purposes and results in the formulation of new education objectives.
32. Evaluation must be continuous and comprehensive to fulfill its diagnostic purpose.
33. A child strives mainly to improve his or her performance and develop potential rather than to compete with others.

34. Teachers accept and respond to the fact that growth patterns will be irregular and will occur in different areas at different times.
35. Individual pupil progress forms are used to record the learning tasks completed, deficiencies that need new assignments to permit mastery, and all other data that will show the child's progress in relation to past achievements and potential or that will help the teacher in suggesting possible future learning experiences for the individual.
36. Evaluating and reporting will consider all five areas of the child's development: aesthetic, physical, intellectual, emotional, and social.

Of course, few nongraded programs have incorporated all of the features identified by Pavan (1972). Those that are most central to the nongrouping concept (and most likely to be implemented in practice under this name) are the ones relating to vertical and horizontal grouping, in particular the abolition of grade levels and of promotion and retention.

#### *Rationales for Nongrading*

The major rationale for a nongraded approach is to provide an alternative to both retention and social promotion (i.e., promoting students regardless of performance). In the view of Goodlad and Anderson (1963) and many who followed them (e.g., Shepard & Smith, 1989), retention is harmful to students, is applied inconsistently, and fails to take into account developmental inconsistencies (e.g., late bloomers), especially among young children. A retained child repeats a whole year of content he or she failed to learn the first time. Spending a year failing to learn a body of curriculum and then spending a second year going over the same curriculum seems to be a poor practice for low achievers. Advocates of nongrading would argue that it is far better to allow such students to move more slowly through material with a high success rate and never have to repeat unlearned content. As noted earlier, nongraded elementary programs use a continuous progress plan, in which a hierarchical curriculum (such as reading or mathematics) is divided into some number of units across the grades, and then students can take as much time as they need to complete the units. A low achiever moving slowly through a continuous progress curriculum may take as many years to reach the fourth grade as a similar low achiever in a graded structure who is retained at some point in grades 1–3, but advocates of nongrading would argue that the continuous progress plan is less stigmatizing, less psychologically damaging, and more instructionally sensible than retention.

Nongraded organization also offers an alternative to traditional forms of ability grouping. Goodlad and Anderson (1963) point out how nongrading can be an improvement on both between-class ability grouping (e.g., high, middle, and low self-contained second grades) and within-class ability grouping (e.g., reading groups). The problem with between-class ability grouping, they argue, is that grouping on any one criterion (such as reading performance or general ability) cannot group students well for any particular skill. For example, a class grouped according to reading skill will have a very broad range of math levels and will even be quite diverse in performance levels on any particular reading task. As a result, the costs of ability grouping in terms of stigmatizing low achievers are not compensated for by any practically meaningful reduction in heterogeneity. Formation of reading groups within heterogeneous classes is similarly flawed in their view. In order to create

homogeneous groups, teachers must have many reading groups, and therefore much class time must be spent on follow-up activities of little instructional value.

In the nongraded plan, students are flexibly grouped for major subjects (especially reading and math) across class and age lines, so that the resulting groups are truly homogeneous on the skills being taught. Further, by creating multiage groups from among all students in contiguous grade levels, it is possible for teachers to create entire reading or math classes at one or, at most, two levels, so that they need not devote much class time to follow-up.

Finally, the nongraded plan is proposed as a solution to the problem of split grades. In many schools with, for example, a class size of 25 and 38 students in each of grades two and three, principals would create one second-grade class, one third-grade class, and one second- and third-grade combination class. In a graded structure, teaching the second- and third-grade class is difficult, as the two portions of the class may be taught completely separately. A nongraded organization, by eliminating the designation of students as second or third graders, solves this problem.

The rationale for the reemergence of the nongraded plan today is similar to that of the 1950s. In the 1980s, retention rates increased dramatically in elementary schools, especially those in large cities (Levine & Eubanks, 1986–1987). This was partly a result of accountability pressures, which focus on the performance of students according to grade level, not age, thereby rewarding districts for such policies as imposing grade-to-grade promotion standards and holding back low achievers (see Allington & McGill-Franzen, 1992; Slavin & Madden, 1991). However, in more recent years, a reaction against high retention rates has taken place, influenced in particular by the work of Shepard and Smith (1989) which documents the negative long-term effects of retention in the elementary grades. Unwilling to return to social promotion (and still under accountability pressures which discourage it), many school districts are currently experimenting with a variety of means of holding standards constant while allowing time spent in the early grades to vary. Among these is the growing use of means of adding a year between kindergarten and second grade for at-risk children—such as, developmental kindergarten, junior kindergarten, transitional first grade, or prefirst programs. However, research on the long-term impacts of these approaches has questioned their value (see Karweit & Wasik, in press). The nongraded primary has been rediscovered as a means of avoiding both retention and social promotion, just as it was in the 1950s.

Another rationale for the nongraded primary school still important today is a reaction against traditional ability grouping. Between-class ability grouping (e.g., high, middle, and low second grades) has been used by a minority of elementary schools, but use of reading groups has been almost universal until very recently (McPartland, Coldiron, & Braddock, 1987). At present, many schools are seeking alternatives to the use of set reading groups (see Barr, 1990), and the nongraded program appears to be a means of doing away with reading groups while still allowing teachers to accommodate instruction to individual needs.

An important factor today in the move toward the nongraded primary that was not a rationale in the 1950s is the trend toward “developmentally appropriate” practices in the early grades. Developmentally appropriate practices are instructional approaches that allow young children to develop skills at their own pace. For example, the National Association for the Education of Young Children (1989) published a position statement, entitled *Appropriate Education in the Primary Grades*, that



described developmentally appropriate education for children ages 5–8. Among the prescriptions were the following:

Each child is viewed as a unique person with an individual pattern and timing of growth. . . . Children are allowed to move at their own pace in acquiring important skills. . . . For example, it is accepted that not every child will learn how to read at age 6, most will learn by 7, and some will need intensive exposure to appropriate literacy experiences to learn to read by 8 or 9. (p. 4)

The NAEYC position paper also supported integrated curriculum and instruction, extensive use of projects and learning stations, cooperative learning, and other strategies quite consistent with the nongraded primary plans of the late 1960s and early 1970s (and with the open classroom of the same period). A book by Katz, Evangelou, and Hartman (1991), published by NAEYC, makes a case for mixed age grouping that emphasizes developmentally appropriate activities and downplays grouping by ability or performance level.

#### *Individually Guided Education*

One important outgrowth of the nongraded concept was Individually Guided Education (IGE), developed and researched by Klausmeier and his colleagues at the University of Wisconsin (Klausmeier, Rossmiller, & Saily, 1977). IGE, in its Wisconsin version or in the one developed by the Kettering Foundation (through I/D/E/A), was a very ambitious, comprehensive restructuring of elementary education. It used a nongraded grouping strategy, in which students were flexibly grouped according to instructional needs rather than age. As in any nongraded elementary school, students could take as much time as they needed to complete the objectives prescribed for each subject. However, IGE affected all aspects of school organization and instruction, not only grouping. Individual plans were prepared for each student, and students were constantly assessed to determine their continuing placements. Instruction could be delivered one-on-one by teachers or peers, to small groups, or (rarely) to large groups. Extensive use was made of learning stations at which students could perform experiments, work on individualized units, or do other individual or small-group activities independently of the teacher. Comprehensive instructional models were developed and implemented in reading, mathematics, social studies, and science. Students were organized into multiage Instruction and Research (I & R) units of 100 to 150 students with (ideally) a unit leader, three to five staff teachers, an aide, and a teacher intern. This team planned and carried out the instruction students received in all subjects. Often, individual teachers would become experts in a given subject and take responsibility for that subject with the entire unit. A building-level Instructional Improvement Committee worked to establish objectives and policies for the school as a whole.

#### **Review Methods**

This review synthesizes the findings of research comparing the achievement effects of nongraded and traditional organizations in the elementary grades (K–6). The review method used is best evidence synthesis (Slavin, 1986), which combines elements of meta-analysis (Glass, McGaw, & Smith, 1981) with those of narrative reviews. Briefly, a best evidence synthesis requires locating all research on a given topic, establishing well-specified criteria of methodological adequacy and germaneness to the topic, and then reviewing this “best evidence” with attention to the

substantive and methodological contributions of each study. Whenever possible, study outcomes are characterized in terms of effect sizes, the difference between the experimental and control means divided by the control group's standard deviation. Details of the review procedures are described in the following sections.

#### *Literature Search Procedures*

Every effort was made to obtain every study ever reported that met the broad substantive inclusion criteria described below. Principal sources included the Education Resources Information Center (ERIC), *Dissertation Abstracts International*, and the reference lists of earlier reviews and of the primary studies themselves. Most of the studies located were doctoral dissertations. These were obtained from University Microfilms International in Michigan or from the Library of Congress in Washington, DC, which maintains microfilm copies of all U.S. dissertations. In a few cases where unpublished documents could not be found or where clarifications were needed about important studies, authors were contacted directly.

#### *Substantive Inclusion Criteria*

Studies were included in an initial search if they could be identified as evaluating nongraded, ungraded, multiage, or Individually Guided Education programs in grades K–6. Studies spanning elementary and middle grades were included, but only data up to grade six were considered.

#### *Methodological Inclusion Criteria*

Studies were included if they met the following methodological criteria, which are identical to those applied in earlier reviews of ability grouping by Slavin (1987, 1990):

1. Some objective measure of achievement was used. Because of their subjective nature, grades were not included as achievement variables. In practice, all achievement outcomes were assessed using standardized measures.
2. Initial comparability of the nongraded and graded samples was established by means of random assignment of students, matching of schools or classes, or matching of individual students within classes or schools. In studies using matching, evidence had to be presented to indicate that either the groups were initially equivalent (within 20% of a standard deviation) or that they were not equivalent in which case pretest data had to be presented to allow for adjustment of posttest scores for pretest differences. Studies that used gain scores or analyses of covariance to control for initial differences between nongraded and graded programs are listed in separate portions of each table, as statistical adjustment for pretest differences cannot be assumed to completely control for their influence on posttests (see Reichart, 1979). Results of these studies should be interpreted cautiously.
3. The nongraded program was in place for at least a semester. All studies located met this standard; in fact, only two studies were less than a year in duration.

Very few studies which used any achievement measure to compare nongraded and graded programs were excluded on the basis of these inclusion criteria. Examples of studies excluded are ones which involved nongraded secondary schools (e.g., Chalfant, 1972); studies without any evidence that nongraded and control groups were initially equivalent and without adjustments for pretests (e.g., Ingram, 1960); and

studies of school organization plans related to, but not the same as, the nongraded program (e.g., Heathers, 1967; Maresh, 1971).

Studies were not excluded if they met the above criteria but failed to present data that would allow for computation of effect sizes. Instead, such studies were discussed in the review and were included in all tables with an indication of the direction and statistical significance of any differences (see below).

### *Computation of Effect Sizes*

Whenever possible, effect sizes were computed in a manner similar to earlier reviews of ability grouping by Slavin (1987, 1990). In general, effect sizes were computed as the difference between the nongraded and graded programs' means divided by the graded program's standard deviation. When reports omitted means or standard deviations, effect sizes were estimated from *ts*, *Fs*, *ps*, or other statistics, using procedures described by Glass, McGaw, and Smith (1981). However, one important departure from the Glass et al. procedures was used when appropriate. If pretest scores were available, posttests were adjusted for them using ANCOVA or raw gain scores. However, denominators in the effect-size computations were always unadjusted individual-level posttest standard deviations. The purpose of these procedures was to avoid situations in which one treatment exceeded another at pretest and posttest to the same degree yet the posttest difference was coded as meaningful. See Slavin (1987) for more on this adjustment procedure and other details of effect size computation.

For some purposes, effect sizes were pooled across studies. Whenever this was done, medians (not means) were computed on all studies from which effect-size estimates could be derived. Pooling effect sizes within well-defined categories of studies can provide a useful summary of the size and direction of effects, but the pooled estimate should always be evaluated in light of the methodological quality and the consistency of results of the individual studies narratively described in the text.

### **Categories of Nongraded Programs**

As noted earlier, nongraded elementary schools have varied widely in their particulars. This variation is not surprising, given that the original conception of the nongraded idea did not pretend to touch on all aspects of school organization and instruction:

Nongrading is a scheme for organizing schools vertically. It does not account for the many problems of organizing schools horizontally. (Goodlad & Anderson, 1963, p. 210)

In looking at studies of nongraded elementary schools over time, an interesting pattern emerges. The earlier studies tended to apply nongrading to only one subject, usually reading. As time went on, studies began to include more than one subject but still to maintain traditional curriculum and instruction; later still, nongraded programs began to incorporate much more radical changes in curriculum and instruction, along with increased use of team teaching, individualized instruction, learning stations, peer tutoring, cooperative learning, and so on. Individually Guided Education (IGE) represented a full flowering of this form of nongrading.

It is possible to distinguish four distinct categories of nongraded programs, and this review considers each type separately. These are as follows:

1. *Nongraded Programs Involving Only One Subject (Joplin-Like Programs)*  
Nine studies, all reported in the 1950s or 1960s, evaluated nongraded plans that only involved one subject. The subject was reading in eight studies, math in one.
2. *Nongraded Programs Involving Multiple Subjects (Comprehensive Programs)*  
Fourteen studies, reported from the late 1950s or 1960s to the early 1980s, evaluated nongraded plans incorporating two or more subjects (and often including all academic subjects). This category adheres most closely to the original conception put forward by Goodlad and Anderson (1963), in that the nongraded programs emphasize continuous progress and flexible, multiage grouping but do not emphasize individualized instruction.
3. *Nongraded Programs Incorporating Individualized Instruction*  
Eleven studies, all but one reported in the brief period from 1969 to 1973, evaluated nongraded programs that emphasized individualized instruction, learning stations, learning activity packages, programmed instruction, and/or tutoring.
4. *Individually Guided Education (IGE)*  
Ten studies evaluated implementations of Individually Guided Education (IGE), described earlier. This was the latest group of studies, with reports appearing over the period from 1972 to 1985.
5. *Studies Lacking an Explicit Description of the Nongraded Program*  
In addition to the four categories discussed above, 12 studies failed to state what was actually implemented in the nongraded programs they evaluated. These were generally *ex post facto* studies, often with large samples, in which the researchers simply accepted principals' words that their schools were nongraded. Given the considerable diversity among implementations that were described, it would be foolish to assume anything about what the independent variable in these studies really was. However, this category is included for the sake of completeness.

### **Research on Nongraded Programs**

The following sections discuss the research on each of the categories of nongraded programs described above. The five sections contain tables summarizing the major characteristics and findings, first of randomized studies, then matched equivalent studies, and finally matched studies lacking evidence of initial equality. Within these categories, the larger studies are listed first. The text usually follows the same order. In general, then, studies listed and discussed earlier in each section can be considered higher in methodological quality than those that come later.

In each table, effect sizes are presented for each measure or subgroup used, and then an overall effect size is presented. Asterisks by effect sizes indicate that the differences were statistically significant, according to the authors. When effect sizes could not be computed, outcomes are characterized as favoring nongraded (+), no difference (0), or favoring graded (−), with asterisks if the differences were significant. A key to all symbols and abbreviations used in all tables appears in an appendix.

### *Joplin-Like Nongraded Programs*

Table 1 summarizes the research on nongraded programs that have as a distinctive feature the homogeneous grouping of students according to performance level in

TABLE 1

*Nongraded programs involving only one subject (Joplin-like programs)*

Article	Grades	Location	Sample size	Duration of program	Design	Test	Effect sizes		Total
							By achievement	By subject	
<u>Randomized studies</u>									
Jones, Moore, & Van Devender (1967)	2-3	Shamokin, Pennsylvania	52 (26 NG, 26 G) (1 school)	1.5 yrs. 3 yrs. (follow-up)	Students and teachers randomly assigned to NG/Joplin or heterogeneous graded classes for reading only.	SAT, LCRT		Rdg. (1.5 yrs.) + .72* Rdg. (3 yrs.) + .33	+ .33
<u>Matched studies with evidence of initial equality</u>									
Halliwell (1963)	1-3	New Jersey	295 (146 NG, 149 G) (1 school)	1 yr.	Compared NG/Joplin in reading and spelling to previous yr. heterogeneous grouping in the graded program. Students had comparable IQ at the beginning of the study.	CAT (gr. 1), MAT (gr. 2-3)		Rdg. + .53*	+ .53*
Skapski (1960)	3	Burlington, Vermont	110 (3 schools)	3 yrs.	Students matched on IQ. Compared NG/Joplin in reading only to heterogeneous grouping in a graded program.	SAT	Su. + .91 Hi. + .48 Av. + .52	Rdg. + .57**	+ .57**
Hart (1962)	4	Hillsboro, Oregon	100 (50 NG, 50 G) (1 school)	3 yrs.	Students matched on sex, mental maturity, age, and SES. Compared NG/Joplin in arithmetic only to heterogeneous grouping in a graded program.	CAT		Math + .46*	+ .46*

Matched studies lacking evidence of initial equality

Bockrath (1958)	4	Archdiocese of St. Louis	(1974 NG, 1622 G) (50 schools)	3 yrs.	Comparison between 1956 students' reading achievement with 1953 students' scores. IQ used to adjust score medians. Stratified sample by size and location of schools.	CAT	Hi. + Lo. +	Rdg. + .51**	+ .51**
Jacquette (1959)	1-6	Grand Junction, Colorado	3517 (1554 NG, 1963 G) (4 schools)	5 yrs.	Schools matched on rdg. achievement and IQ. Pretest used to compute gain scores.	CAT, GPRT		Rdg. + .03	+ .03
Moore (1963)	1-2	Wayne, Michigan	621 (292 NG, 329 G) (4 schools)	1 yr.	Schools matched on SES and curriculum. Change scores used to control pretest achievement significant differences. Compared NG/Joplin in reading only to conventional graded plans.	MAT	Hi. - .22 Av. - .43** Lo. - .29	Rdg. - .41**	- .41**
Enevoldsen (1961)	1-3	Lincoln, Nebraska	420 (210 NG, 210 G) (7 schools)	2 yrs. (2 sch.) 3 yrs. (1 sch.)	Schools matched on SES. IQ used as a covariate. Compared NG/Joplin in reading only to graded programs.	CAT		Rdg. 0	0
Kierstead (1963)	3-8	Orwell, Vermont	277 (111 NG, 166 G) (2 schools)	1 yr.	Students equated and classified by IQ and pretest. Pretest used to compute gain scores. Compared NG/Joplin in reading only to ability grouping in a graded plan.	ITBS	Hi. - .01 GE Av. + .08 GE Lo. - .14 GE	Rdg. - .02 GE	- .02 GE

Note. A key to the abbreviations used is provided as an appendix to this article.

\* $p < .05$

\*\* $p < .01$

only one subject. These plans can be labeled Joplin-like programs because they share with the Joplin Plan the idea of regrouping students for just one subject (usually reading), ignoring grade levels or ages. Nine studies, all done during the 1960s, are included in this category. Most of them were described under the Joplin Plan arrangement in an earlier synthesis of ability grouping and student achievement in elementary schools (Slavin, 1987). These studies appeared early in the nongraded movement, suggesting that the earlier implementations were more conservative (affecting only one subject) than those which appeared later.

Results from five of the nine studies found strong positive effects for the nongraded plans, three studies reported no differences between them and graded plans, and one significantly favored the graded program.

Jones, Moore, and Van Devender (1967) randomly assigned students and teachers to nongraded or traditional classes. Students in the nongraded classes were assigned to heterogeneous classes but regrouped across grade levels for reading. They proceeded through nine reading levels and were continually regrouped on the basis of their reading performance. Within each reading class, teachers had multiple reading groups and used traditional basal readers and instructional methods. The results of this study supported the efficacy of the nongraded program. After 3 semesters, reading scores for experimental students on three standardized scales were considerably higher than for control students ( $ES = +.72$ , or about  $+.41$  grade equivalents). After 3 years in the program, experimental-control differences had diminished but were still moderately positive ( $ES = +.33$ ).

Three studies compared Joplin-like nongraded classes to matched control classes and presented evidence of initial comparability. In the largest of these, Halliwell (1963) evaluated a nongraded primary that was virtually identical to the Joplin Plan. Students in first through third grades were regrouped for reading only and remained in heterogeneous classes the rest of the day. Spelling was also included in the regrouped classes for second and third graders. The article was unclear as to whether within-class grouping was used in regrouped reading classes, but there was some indication that reading groups were not used. Results indicated considerably higher reading achievement in nongraded classes than in the same school the year before nongrading was introduced ( $ES = +.53$ ). Scores were higher for nongraded students at every grade level, but by far the largest differences were for first graders, who exceeded earlier first grade classes by  $+.94$  grade equivalents ( $ES = +1.22$ ). It is important to note that mathematics achievement, measured at the second- and third-grade levels, also increased significantly more in the nongraded classes than in previous years ( $ES = +.51$ ). Because mathematics was not part of the nongraded program, this finding suggests the possibility that factors other than the nongraded plan might account for the increases in student academic achievement. However, the author notes that teachers claimed to have been able to devote more time to mathematics because the nongraded program required less time for reading, spelling, and language instruction than they had spent on these subjects in previous years.

A study by Skapski (1960) also evaluated the use of a Joplin-like nongraded organization for reading only. The details of the nongraded program were not clearly described, but it appears that reading groups were not used in regrouped classes and that curricula and teaching methods were traditional. Two comparisons were made. First, the reading scores of students in the nongraded program were compared to the same students' arithmetic scores, on the assumption that because arithmetic was not

involved in the nongraded plan any differences would reflect an effect of nongrading. Results of this comparison indicated that second- and third-grade-aged students achieved an average of 1.1 grade equivalents higher in reading than in arithmetic. Further, scores of third graders who had spent 3 years in the nongraded program were compared to students in two control schools matched on IQ. Results indicated that the nongraded students achieved at a much higher level in reading than did control students ( $ES = +.57$ ) but that there were no differences in arithmetic (which was not involved in the nongraded program). Differences were particularly large for students with IQs of 125 or higher ( $ES = +.91$ ) but were still quite substantial for students with IQs in the range 88–112 ( $ES = +.52$ ).

Only one study evaluated the use of a Joplin-like program in mathematics. This was a study by Hart (1962). Experimental students were regrouped for arithmetic instruction across grade lines and were taught as a whole class. Students were frequently assessed on arithmetic skills and reassigned to different classes if their performance indicated that a different level of instruction was needed. Experimental students who had spent 3 years in the nongraded arithmetic program were matched on IQ, age, and SES with students in similar schools using traditional methods. It was not stated whether control classes used within-class ability grouping for arithmetic instruction. Results indicated an advantage of about one half of a grade equivalent for the experimental group ( $ES = +.46$ ).

Five studies matched Joplin-like nongraded classes with graded ones and dealt statistically with initial differences among students. In a study in Catholic schools in the Archdiocese of St. Louis, Bockrath (1958) analyzed the largest sample. She conducted three studies in one: first, a comparison of the fourth-grade reading test scores of 1953 and 1956 in the 366 schools of the archdiocese (12,450 students); second, the same comparison for a stratified sample of 50 of these schools (3,596 students); and third, a 3-year study of one of the archdiocesan schools to examine how the nongraded primary functioned (106 students). In 1956, students had been in the nongraded program for 3 years, while the 1953 pupils had received graded instruction. Besides a 1 point difference in mean IQ, the students differed in entrance age. An effort was made by the author to adjust the fourth-grade reading score medians in relation to IQ, but only a narrative description took into account the new entrance age adopted for the second group of students (a mean increase of 2 months for first-year primary entrants in 1953). The results clearly favored the nongraded plan ( $ES = +.51$ ), which was characterized by the creation of flexible skill-level groupings.

The only study of a Joplin-like program that found a clear advantage for a graded plan was presented by Moore (1963). First- and second-grade students' reading and arithmetic achievement scores were compared for nongraded and conventionally graded schools. Following comparable instructional practices, second-grade students' reading scores in the graded and nongraded schools did not differ significantly ( $ES = -.12$ ). However, a substantial negative difference (i.e., favoring the graded program) was found among first graders' reading scores ( $ES = -.70$ ). As the author surveyed teachers with respect to which reading material their pupils were using, he found that first-grade control students were reading approximately one basal reading text higher than the experimental group. Arithmetic scores did not differ for first- and second-grade students, but no description is given of the instruction in this subject.

Enevoldsen (1961) did another study that found schools which differed in label rather than in organizational structure. The graded and nongraded schools chosen



from the same public system had similar nongraded reading programs. Consequently, no significant difference was found in students' reading achievement. Two additional studies of relatively low quality, by Jacquette (1959) and Kierstead (1963), reported effect sizes that were close to zero. Both studies stated that learning levels established in the nongraded programs followed very closely the sequential pattern of graded reading skills, and as a result few differences in outcomes were found.

*Summary and discussion: Joplin-like nongraded programs.* Overall, the findings of methodologically adequate studies of this type of nongraded program were consistent. All studies exhibiting good methodological quality (randomized and matched studies with evidence of initial equality) found substantial positive results in favor of the nongraded program. The median effect size for the four best quality studies was + .50; for all seven studies from which effect sizes could be estimated, it was + .46. The matched studies lacking evidence of initial equality that do not report positive results were characterized by similar reading programs; the biggest difference between them appeared to be their label. Two features were important in almost all of the successful nongraded programs evaluated: flexibility in pupil grouping, with frequent assessment of mastery at each level; and increased amounts of teaching time for the homogeneous instructional groups. Because each teacher had to manage fewer groups, there was less need for independent follow-up activities, such as worksheets in reading. Perhaps this last characteristic is one of the most important elements that favors students in a nongraded program: More homogeneous groups allow teachers to define more specific objectives for instruction, and children receive a greater amount of direct teaching.

#### *Comprehensive Nongraded Programs*

Some studies described plans in which more than one subject was nongraded. These 14 studies, summarized in Table 2, were conducted from the late 1960s to the beginning of the 1980s. Only three evaluations presented small (and nonsignificant) negative total effect sizes, while 8 of the 10 that presented results favoring the nongraded plans reported statistically significant differences.

Among the eight studies with evidence of initial equality, Brody's (1970) had the largest sample size. It evaluated a nongraded program in which first and second graders had to pass a series of sequential steps in several subjects at 90% mastery and were placed in groups according to their mastery of specific skills (regardless of grade level). Vertical advancement of students was strongly emphasized. At the time of assessment, first graders had been in this program 1 year, and second graders 2 years. Both groups of nongraded students gained significantly more than did students in graded classes ( $ES = +.20$ ). Effects were particularly large in mathematics ( $ES = +.73$ ). This study was somewhat flawed by the fact that, before matching, the nongraded students were 5.4 points higher in IQ than their graded counterparts.

The only matched equivalent study to find no differences in achievement between nongraded and graded programs was done by Otto (1969), in a laboratory school at the University of Texas. Unlike many other studies of nongraded plans, this study fully described the nongraded intervention, designed to be a full-scale implementation of the Goodlad and Anderson (1963) nongraded plan. Unfortunately, experimental and control groups did not differ on many elements held to be essential to the nongraded program. Teachers of the nongraded classes did assign students to instructional groups across grade lines, had students use materials suited to different levels,

and provided less whole-class instruction than did teachers in the graded program. However, the nongraded classes did not use more subgroups than graded classes and did not reduce the heterogeneity of subgroups. Because the experiment took place in a laboratory school, it may be that control classes were of high quality and control teachers may have used many aspects of nongrading in their classes.

It is interesting to note that two other studies conducted in university laboratory schools by Muck (1966) and Ross (1967) also failed to find differences between nongraded and graded classes. The principal differences between the graded and nongraded programs in the Muck study concern the sequence in which the curriculum was taught and the policy that children in the nongraded program remain with the same teacher for 3 years while three different teachers faced students in the graded plan. These are not key issues in the usual characterization of the nongraded approach.

Perrin (1969) also found slight differences in favor of the nongraded plans ( $ES = + .11$ ). As he analyzed the data at the end of each year in a 3-year study, it became clear that, as time passed, the results started to differ significantly. Perrin evaluated a nongraded program in which a minimum skills chart was used to trace the progress of each child. These basic skills in reading, language, and arithmetic were divided into levels, and children moved through them at their own paces.

Buffie (1962) compared the scores children obtained during their last year in a graded or nongraded primary program. In the graded plan, pupils worked on the same program at the same time in all subjects except reading. In the nongraded plan, grouping within as well as between classrooms was done in reading, arithmetic, and spelling. Team teaching practiced in the nongraded schools also differentiated the plans. The results favor the nongraded plan (total  $ES = + .34$ ), especially on the language subtest ( $ES = + .67$ ).

Another study that pointed out sharp differences in the instructional practices of the two groups compared was done by Guarino (1982). Using an index for non-gradedness (Pavan, 1972), he tested the congruence between labels and structures. The main distinction was that grouping in the nongraded program was intended to provide an appropriate level of instruction for all students and was guided by frequently administered diagnostic tests to discover deficiencies in skill areas. High and low achievers in the graded program had problems in receiving instruction appropriate to their special needs. The standardized scores differed significantly in favor of the nongraded program (total  $ES = + .34$ ), especially in the reading subtests ( $ES = + .49$ ) and for the older students.

Ramayya (1972) reported positive results for the sixth-year students in a non-graded school (total  $ES = + .42$ ). For 6 years, these students attended reading and arithmetic classes that were reorganized into several levels. This study confirms the findings reported by Perrin (1969), Brody (1970), Buffie (1962), and Guarino (1982): The longer the duration of the nongraded program, the greater its favorable impact on student academic achievement.

Among the matched studies lacking evidence of initial equality, the largest was a study by Zerby (1960). Instructional practices were similar in the two programs he compared. Reading and arithmetic texts differed between programs, and the non-graded program provided the students with the same teacher for all 3 years, while different teachers every year instructed children in the graded plan. Despite the

TABLE 2

*Nongraded programs involving multiple subjects (comprehensive programs)*

Article	Grades	Location	Sample size	Duration of program	Design	Test	Effect sizes		Total
							By achievement	By subject	
<u>Matched studies with evidence of initial equality</u>									
Brody (1970)	1-2	Pennsylvania	603 (362 NG, 241 G) (3 schools)	2 yrs.	Students matched on IQ.	SAT	Hi. + Lo. +	Rdg. +.20 Math +.73**	+.46**
Otto (1969)	3-5	Austin, Texas	450 (2 upper middle-class lab schools)	2 yrs.	Students matched on pretest achievement.	MAT, ITBS		Rdg. 0 Math 0	0
Perrin (1969)	1-3	Little Rock, Arkansas	288 (144 NG, 144 G) (13 schools)	3 yrs.	Schools matched on SES. Students matched on IQ, age, sex, and race.	MAT (1-2), ITBS (3)		Rdg. +.08 Math +.14	+.11
Buffie (1962)	3	Cedar Falls, Iowa	234 (117 NG, 117 G) (8 schools)	3 yrs.	Schools matched on SES, enrollment, class size, and teachers' experience. Students matched on sex, age, and intelligence.	ITBS	Hi. +.39 Lo. +.19	Rdg. +.19 Math +.17 Lang. +.67**	+.34**
Guarino (1982)	2-5	New Jersey	162 (81 NG, 81 G) (2 schools)	5 yrs.	Schools matched on SES and ethnic mix. Students matched on age, sex, and IQ.	CAT		Rdg. +.49** Math +.19	+.34*
Ramaya (1972)	6	Darmouth, Nova Scotia	160 (80 NG, 80 G)	6 yrs.	Students matched on sex, IQ, and SES.	TBS		Rdg. +.41* Math +.25 Lang. +.59*	+.42*
Muck (1966)	1-3	Buffalo, New York	148 (1 lab school)	3 yrs.	Students matched on mental maturity.	MAT, ITBS		Rdg. +.04 Math -.36 Lang. +.15	-.06
Machiele (1965)	1	Urbana, Illinois	100 (50 NG, 50 G) (1 school)	1 yr.	Students matched on IQ, mental age, and chronological age. Compared students in NG program to students in previous yr.	CAT		Rdg. +.61** Math +.38	+.49*

Matched studies lacking evidence of initial equality

Zerby (1960)	3	Morristown, Pennsylvania	394 (187 NG, 207 G) (2 schools)	3 yrs.	Schools matched on SES. IQ score used to compute achievement beyond anticipated achievement level.	CAT	Rdg. +.10 Math +.57**	+.34*
Chastain (1961)	4-6	Rangey, Colorado	360 (120 NG, 120 G) (1 school)	1 yr.	Students matched on sex and IQ. Pretest used as a covariate.	MAT	Rdg. +.01 Math -.09	-.04
Lawson (1973)	1, 3, 5	Kokomo, Indiana	338 (6 schools)	1, 3, & 5 yrs.	IQ used as a covariate.	CAT	Rdg. +**	+**
Ross (1967)	1-3	Bloomington, Indiana	314 (128 NG, 186 G) (1 lab school)	6 months	Pretest and IQ used as covariates. Students nonrandomly assigned to NG and G programs in the school.	MAT	Rdg. +.06 GE Math +.06 GE	+.06 GE
Morris (1968)	1-3, 5	Montgomery County, Pennsylvania	117 (57 NG, 60 G) (1 school)	3 yrs.	IQ used as a covariate. Compared students in NG program to students in previous year. Intervention stopped after 3 years.	ITBS, SAT	After 3 yrs. +* After 5 yrs. +**	+**
Gumpper, Meyer, & Kaufman (1971)	1-4	Pennsylvania	(2 schools)	1 yr.	Schools matched on SES, enrollment, teachers' characteristics, and students' previous academic achievement. Pretest used to compute gain scores.	DLRT, SAT	Gr. 1 Rdg. 0 Math (+) Gr. 2-4 Rdg. (-) Math -** Lang. -**	(-)

*Note.* A key to the abbreviations used is provided as an appendix to this article.

\* $p < .05$

\*\* $p < .01$

resemblance in instructional practices, the results significantly favored the non-graded program (total ES = +.34), especially in arithmetic (ES = +.57).

Lawson (1973) and Morris (1968) conducted studies that had several similarities. Reading and mathematics programs were organized by levels, and regrouping allowed teachers to teach classes of students all at one level in each subject. In the nongraded plans studied, team teaching was described by Lawson, while Morris emphasized the fact that teachers did not face more than two different ability groups per class. Both studies found positive results that increased with time. After 3 and 5 years, significant differences favored the nongraded programs.

After only 1 year of intervention, Chastain (1961) evaluated the academic achievement of students in an intermediate school that shifted from a graded structure to achievement-level grouping in reading classes and finally to a nongraded plan. No differences were found in the reading achievement of students belonging to either plan; a negative difference in arithmetic achievement became smaller in the second year of homogeneous grouping (first year of the nongraded plan).

Another study that evaluated what could be considered a comprehensive non-graded program was conducted by Gumper, Meyer, and Kaufman (1971). Test scores from children attending the first 4 years of school in a continuous-progress program and in a modified self-contained graded program were compared. Some ability grouping was used for mathematics and reading classes at the same grade level in the control school. Students changed classes several times during each day, breaking some of the atmosphere of the self-contained classroom situation. In the experimental school, children were grouped homogeneously according to achievement for language arts and mathematics. When the nongraded program was introduced, problems with ability grouping occurred, and teachers had to deal with as many as three different levels of children at the same time. The fact that the posttest was administered at the end of the first year of the nongraded program was Gumper's main explanation for the positive results for first graders and the negative results for second-, third-, and fourth-grade students. Rather than a true difference between graded and nongraded plans, Gumper believes that the lower achievement gains were more a function of problems of reorientation for older students in the continuous-progress school.

*Summary and discussion: Comprehensive nongraded programs.* Findings from this group of studies consistently favored the nongraded program. Almost all of its positive results were significant; not one study found significant differences in favor of the graded plan. The median effect size for the matched equivalent studies was +.34, and it was the same for all nine studies from which effect sizes could be estimated. Among those studies that did not report any significant difference, three were conducted in university laboratory schools, and another three found equivalence in the first year of the program but started to see favorable changes in subsequent years. In the case of laboratory schools, control classes were similar to experimental ones, or they appeared to be very high quality classes. Perhaps for these reasons, significant differences did not appear in those circumstances. Across many studies, greater duration of the program was associated with higher positive differences. Other common characteristics of academically successful nongraded plans were subjects organized by levels, use of texts written in accordance with those levels, and regrouping of students in multiage environments that allowed teachers to reduce the heterogeneity of their instructional groups.

*Nongraded Programs Incorporating Individualized Instruction*

Many studies of nongraded programs included indications that individualized instruction was an important part of the nongraded plan. These individualized approaches included one-to-one tutoring, programmed instruction, and learning activity packages. Two examples of the types of individualization adopted are as follows:

Most students would be on contracts of work . . . [that] might last from one to five days with the student coming to the teachers only in particular moments of difficulty. (Bowman, 1971, p. 46)

The Individually Prescribed Instruction mathematics program . . . was used in the model school. This individualized system of instruction provided each student with the opportunity to work on undeveloped skills, to obtain a diagnosis of new learnings, and to receive a prescription for the next sequence of material to be mastered. Math specialists, instructional aides, and volunteer aides were available to pupils on a one-to-one basis. (Jeffreys, 1970, p. 30)

All but one of the 11 studies of this type were published in a brief period from 1969 to 1973, with a median of 1971. This is considerably later than the time frame in which the studies summarized in Table 1 appeared. The median publication date for the Joplin-like programs is 1962, and, for the comprehensive (multiple subjects) programs, it is 1969. What this progression suggests is that individualized instruction increasingly became part of the nongraded elementary school in the late 1960s and early 1970s at a time when individualization was gaining popularity in North American schools in general.

Table 3 summarizes the characteristics and findings of the 11 studies of nongraded programs including individualized instruction. Only one of these, conducted by Higgins (1980), randomly assigned students to nongraded or graded classes. In that study, reading was the only subject of interest. The physical arrangements and instructional practices in both settings were essentially different. For the most part, instruction for children in the nongraded classrooms was individualized. The graded classes were grouped by ability, and discussions took place in each group to check comprehension. The scores obtained by each group of students on the Metropolitan Reading Test did not show any significant difference ( $ES = +.02$ ).

Sie (1969) studied pupils from second, third, and fourth grades who were in two schools, one with a traditional graded plan and the other with a nongraded plan. The students were matched according to their SAT scores. Both schools shared similar group instruction in the areas of work study skills, social studies, and science. The principal difference between them was that the nongraded school emphasized individualized instruction in reading and arithmetic, while the graded one used some form of within-class grouping for reading. Of 24 gain scores computed for the SAT subtests, the nongraded students scored significantly better in one, arithmetic computation, while the graded students performed significantly better on paragraph meaning and language subtests. The overall effect size was near zero.

Jeffreys (1970) evaluated the academic achievement of children in a nongraded program characterized by an open-space building design, team teaching, and individualized instruction with that of children in a traditional, graded plan. In the nongraded school, reading ability levels were used to group pupils for language arts, and skill levels were used to group them homogeneously in science. Students were

**TABLE 3**  
*Nongraded programs incorporating individualized instruction*

Article	Grades	Location	Sample size	Duration of program	Design	Test	Effect sizes		Total
							By achievement	By subject	
<u>Randomized studies</u>									
Higgins (1980)	3-5	Baton Rouge, Louisiana	246 (75 NG, 171 G) (3 schools)	1 yr.	Students randomly assigned to nongraded/combination or traditional reading classes. Pretest used to compute gain scores.	MRT	Hi. (+) Lo. (+)	Rdg. +.02	+.02
<u>Matched studies with evidence of initial equality</u>									
Sie (1969)	2-4	Ames, Iowa	124 (67 NG, 67 G) (2 schools)	1 yr.	Schools matched on SES. Students matched on SAT scores. Pretest used to compute gain scores.	SAT		Rdg. +.03 Math +.14 Lang. -.11	+.02
Jeffreys (1970)	3, 5	Howard County, Maryland	88 (44 NG, 44 G) (2 schools)	1 yr.	Schools matched on SES. Students matched on pretest achievement measure. Pretest scores and parent occup. status used as covariates.	ITBS		Rdg. +.08 Math -.13	-.03
Wilt (1970)	4	Chicago suburb, Illinois	84 (32 NG, 52 G) (2 schools)	4 yrs.	Students matched on IQ and age.	ITBS		Rdg. +.49 GE* Math +.10 GE Lang. -.27 GE	+.11 GE
<u>Matched studies lacking evidence of initial equality</u>									
Ward (1969)	1-2	Fort Worth, Texas	797 (376 NG, 421 G) (4 schools)	2 yrs.	Schools matched on SES, race, and available resources. IQ, age, and readiness scores used as covariates.	MAT		Rdg. + Math (+)	+
Burchyett (1972)	3-5	Grand Blanc, Michigan	535 (332 NG, 203 G) (2 schools)	2 yrs.	Schools matched on SES. Pretest used as a covariate.	STEP		Rdg. -.06 Math -.10	-.08

Bowman (1971)	1-6	Burlington, North Carolina	457 (313 NG, 144 G) (2 schools)	1 yr.	IQ used as a covariate. Pretest used to compute gain scores.	MAT		Rdg. +.27* Math +.28*	+.28*
Case (1970)	5	Montgomery County, Maryland	269 (131 NG, 138 G) (4 schools)	1 yr.	Schools matched on SES. Students matched on age, sex, race, and SES (higher IQ scores for control group). Pretest used to compute gain scores.	SAT	Hi. +.18 Av. +.14 Lo. -.01	Rdg. +.01 Math +.16	+.09
Killough (1971)	1-8	Houston, Texas	267 (132 NG, 135 G) (4 schools)	3 yrs.	Schools matched on SES and ethnic distribution. IQ used as a covariate. Pretest used to compute gain scores.	SRAAS		Rdg. +* Math +*	+*
Givens (1972)	5	Saint Louis, Missouri	100 (50 NG, 50 G) (1 lab, 1 control school)	1 yr.	Students randomly selected from two populations of students that received either individualized or traditional instruction. Ex post facto experimental design. Pretest used to compute gain scores.	ITBS	Hi. (+) Av. 0 Lo. (+)	Rdg. -.11 Math +.10	.00
Walker (1973)	1-12	Kentucky	96 (32 NG, 64 G)	12 yrs.	Schools rated on an eight dimension scale, the Nongradedness Assessment Scale. Longitudinal study to determine the long term effects of NG and G primary school years (1-3). Rate of progress used as a covariate.	CAT		Rdg. +.24 Math +.14 Lang. +.17	+.18
Snake River School District (1972)	1-3	Blackfoot, Idaho	78 (39 NG, 39 G) (2 schools)	1 yr.	Students matched on SES. Pretest used to compute gain scores.	SRAAS		Rdg. .00 Math .00	.00

*Note.* A key to the abbreviations used is provided as an appendix to this article.

\* $p < .05$

\*\* $p < .01$



grouped heterogeneously for social studies and health. In addition to spending more time in individualized and small group settings (math and spelling instruction followed an individualized system), nongraded students were found to initiate verbal interaction with teachers and to be involved in after-school activities a significantly greater number of times. However, no significant differences were found on the Iowa Test of Basic Skills.

Another evaluation of a nongraded program operating in an open area and using team teaching was done by Wilt (1970). The author administered a teacher questionnaire to identify differences between both schools studied with respect to their internal structure, operating procedures, and teacher and student flexibility. Teachers in both schools supported the basic concept of individualized instruction, and it appears that those in the nongraded program used it somewhat more. There is no mention of its use in any specific subject; apparently, it was used whenever the need arose. Consequently, criteria for grouping in the nongraded plan were more diverse (interest, academic achievement, student-teacher relationship) than in the graded program, where homogeneous grouping according to performance level prevailed. Despite the differences in vertical and horizontal organization, students' scores on the Iowa Test of Basic Skills did not differ significantly.

Among the matched studies lacking evidence of initial equality, Ward (1969) investigated the largest sample. He compared the academic achievement of children in four different schools, two of them implementing nongraded programs and two following graded plans. Results favored the experimental group in each of the 72 comparisons (although only 16 of these were statistically significant differences). Ward notes that the nongraded schools differed mainly in the larger amount of time used by their teachers and pupils in reading and arithmetic. The nongraded approach exhibited a more flexible use of time and provided "the kind of 'atmosphere' which is conducive to the individualization of instruction" (Ward, 1969, p. 168).

Burchyett (1972) found the largest differences in favor of the conventional school among the studies reporting use of individualized instruction (total ES =  $-.08$ ). None of these, however, were statistically significant for all of the three grades studied. He compared children attending a nongraded, multiaged, team-taught school with children attending self-contained classrooms in a graded organization. Unfortunately, the author did not specify which areas of instruction were approached on an individual basis, which were characterized by multiage grouping, and how organizational patterns differed in the schools under study.

One of the two studies in Table 3 reporting significant differences in favor of the nongraded program was a study by Bowman (1971). He compared pupils from first to sixth grade in a conventional graded school with students in a nongraded program from another school. The latter used individualized instruction, team teaching, flexible grouping, and learning centers. Individual work was emphasized in reading and mathematics; contracts for work on individualized units were agreed on by teachers and pupils. Grouping across class and grade lines was the organizational arrangement for all other subjects, although curriculum changes were also undertaken for social studies, music, and art (science was an exception due to the lack of time to plan new units adaptable to a multiage situation). Strong positive effects on the academic achievement of intermediate students were found (ES =  $+.52$ ), but the nongraded plan did not have similar effects on the academic achievement of primary students (ES =  $+.06$ ).

Killough (1971) reported another study with significant positive effects of a nongraded plan implemented in an open-space school. In this study, children benefited from being in a nongraded program from first grade through the junior high school years. After 3 years in the program, pupils had significantly higher cognitive achievement gains than control students. Details of the intervention were not described.

Walker (1973) also studied the long-term effects of graded and nongraded primary programs. After rating the degree of nongradedness of each of the six programs he studied, his conclusions after comparing those that truly followed the nongraded principles with all the other plans resembled those of Bowman (1971). Scores began to differ significantly when children remained in a nongraded school after the primary unit. Walker found the greatest gap in academic achievement favoring these children at the fifth-grade level. From his descriptions of the six programs studied, they differed mainly in the graded materials and terminology used in four schools and absent in two. Individualized instruction and grouping across grades were present in most schools, although with varying emphasis in reading and mathematics.

A sharper contrast between self-contained classes and nongradedness with individualized instruction was evident in a study by Case (1970). It compared three conventional elementary schools to one nongraded middle school. The elementary schools used ability grouping and concentrated primarily on the development of reading and mathematics skills (instruction was given to smaller homogeneous groups), while pupils in the nongraded program were encouraged to do independent study in these subjects. The study reported no significant differences between the gain scores of each group of students (total ES = +.09).

Givens (1972) evaluated fifth graders in two different schools. The demonstration school featured team teaching, multiage grouping of pupils, an open-space concept, and individualization of instruction. Local universities contributed interns and student teachers to the schools' staff. No further description of the differences in instructional practices was provided. The standardized test scores did not favor any of the schools (total ES = .00). A zero effect size was also obtained in a small study conducted in the Snake River (Idaho) School District (1972).

*Summary and discussion: Nongraded programs incorporating individualized instruction.* Considered together, the results of research on these nongraded programs were remarkably consistent. No significant differences appeared in most studies. A median effect size of essentially zero (ES = +.02) was found across the nine studies from which effect sizes could be computed. These findings suggest that nongraded programs using individualized instruction were equivalent to graded plans in terms of academic achievement. As the nongraded plans became more complicated in their grouping arrangements, they apparently lost the comparative advantage of Joplin-like or comprehensive nongraded programs.

There is one interesting trend in the data on nongraded programs using individualized instruction: More positive effects were obtained with older than with younger children. It may be that students need a certain level of maturity or self-organizational skills to profit from a continuous-progress program that includes a good deal of independent work. Another indication of this is the observation that the longer the duration of the program, the better the results.

#### *Individually Guided Education (IGE) Programs*

Ten studies, most of which were done in the 1970s, met the inclusion criteria for this review. All of these evaluated the University of Wisconsin IGE program, not the

Kettering model (Klausmeier, Quilling, Sorenson, Way, & Glasrud, 1971). Since nongradedness is a characteristic of IGE schools, comparisons between them and non-IGE schools cast light on the effects of programs with a stress on individualized instruction. Although the degree of implementation of IGE processes varied from one research setting to another, IGE concepts, components, and practices were clearly established by its developers at the University of Wisconsin. As an ideal nongraded plan, the IGE program takes into account individual differences and uses specialized curriculum materials in reading, mathematics, and other subjects. But the IGE program is far more complicated than a usual nongraded program. Most reports do not provide any description of the type of intervention actually experienced by the experimental schools; it is implied that their organization follows the structure set in the implementation guidelines of 1971.

Schneiderhan (1973) did the only randomized study of IGE programs. She compared two experimental programs, Individually Guided Instruction (IGI) and Individually Guided Education (IGE), with a traditionally graded program. Fourth-through sixth-grade students from one school were randomly selected to follow an IGI or a traditional program. Their ITBS scores and the scores from a third group of students in a second school with an IGE program were used in the comparison. Both IGE and IGI programs were characterized by nongradedness, team teaching, multi-age grouping, differentiated staffing, and open-space environments. It can be inferred from the short descriptions provided that the only difference between them was the additional components an IGE program has: an Instructional Improvement Committee and a Systemwide Program Committee. There were no consistent significant differences between the academic achievement of the three groups. It is possible that only 1 year in the IGI program after several years (at least 3) in the conventional graded plan could not dramatically affect children's performance.

Price (1977), in the largest matched study lacking evidence of initial comparability, started by measuring the level of implementation of the 35 processes emphasized in IGE schools. Then he compared students' academic achievement in schools that were high implementers of the IGE philosophy with students' achievement in schools that were low implementers. Schools identified as high implementers, compared to low implementers, were associated with either higher or comparable pupil outcomes. Significant differences were found for reading achievement, but these cannot be associated with specific classroom practices and organization because no description of them was provided. The findings of this study are biased by the fact that high implementers are probably better schools than low implementers, and this superiority is likely to be related to student achievement differences.

Five other studies showed no consistent differences among the schools compared. Except for Kuhlman's (1985), none of them carefully described the organizational arrangements adopted by each program. Biernacki (1976) evaluated an IGE multi-unit program implemented by an inner city elementary school. The academic achievement of selected students attending grades three through six from the multi-unit school and from a self-contained graded school was compared, and no significant differences were found (total  $ES = +.17$ ). Klaus (1981) tried not only to compare achievement of students in IGE and non-IGE schools but also to determine the effect of the number of years spent by a student in each program on his or her performance. His sample consisted of students who attended the same elementary schools and remained in the system through 11th grade. The IGE program ended at the sixth-

grade level. Overall student change scores from fourth to sixth grade showed no significant effects of the program (total ES = +.05), and no significant differences were found in the achievement of 11th-grade students. This study reported larger differences favoring the scores of fourth graders in IGE schools. These differences became smaller in subsequent grades.

In the only study that favors non-IGE schools, Flowers (1977) studied students' academic achievement in open-space IGE schools and in traditional schools. He found no significant differences between their standardized test scores (total ES = -.25). Also, Henn's (1974) comparison of IGE and non-IGE programs did not yield significant differences in the academic achievement of students in the two programs (total ES = +.03).

Kuhlman (1985) compared four types of school organization: conventional graded schools, traditional alternative schools (also graded), open alternative schools, and IGE schools. From the author's descriptions, it is clear that the last two types of schools emphasized individualized instruction and used a nongraded approach. In regression analyses, the variable for school organization did not yield a significant coefficient.

Three more studies, besides the one conducted by Price (1977), found significant differences in favor of IGE schools. Bradford (1972) studied an IGE school characterized by multilevel programs for reading, mathematics, and spelling by an effective use of a multimedia approach and community resources for classroom enrichment and by experiences with dramatization as a classroom technique. Preassessment tools were used to assemble students in small groups, in an independent study program, or in a one-to-one relationship. Students' gains were close to statistical significance in reading and significantly greater in mathematics in the IGE school than in self-contained classrooms.

Burtley's (1974) comparison of IGE and traditional programs favored the former with respect to academic achievement (total ES = +.48). The author reported substantial differences in teachers' behavior and in instructional programs between the two plans compared. Teachers in the IGE school assumed more professional responsibilities and engaged in team teaching. Children in this school were exposed to several instructional modes: large or small multiage groups, one-to-one tutoring, independent study, and study in dyads. Another important difference was that an effort was made to avoid teaching subjects in isolation from each other; consequently, skills were frequently reinforced in the IGE program.

Finally, a small study done by Soumokil (1977) reported the largest total effect size (+.80). As in the Price (1977) study, the research started by assessing the differences between an IGE elementary program in one school with a standard program in another. After confirming the character of each school's label, the author proceeded to compare students' standardized test scores. IGE students scored significantly higher.

*Summary and discussion: IGE programs.* Overall, research findings on IGE schools resemble results obtained by other studies on nongraded programs incorporating individualized instruction (Table 3). The median effect size across six studies from which effect sizes could be computed was near zero (ES = +.11). Nevertheless, four studies reported significant differences in favor of IGE schools, and all of these were evaluations of schools that clearly differ from one another. Schools closer to a full implementation of IGE concepts seemed to supply students

**TABLE 4**  
*Individually Guided Education (IGE)*

Article	Grades	Location	Sample size	Duration of program	Design	Test	Effect sizes		Total
<u>Randomized studies</u>									
Schneiderhan (1973)	4-6	Roseville, Minnesota	484 (206 IGE, 88 IGI, 190 G) (2 schools)	1 yr.	Students randomly selected to individually guided instruction or traditional programs in the same school. IQ and pretest scores used as covariates.	ITBS	By achievement (IGI & control)	By subject Rdg. 0 Math 0 Lang. 0	0
							(IGE & IGI & control)	Rdg. (-) Math 0 Lang. (+)	
<u>Matched studies lacking evidence of initial equality</u>									
Price (1977)	4, 6	Iowa	1081 (637 Hi., 444 Lo.) (14 schools)	3 yrs.	Comparison of high and low implementers of 35 processes employed in IGE. Schools matched on size, SES, and location. IQ used as a covariate.	ITBS		Rdg. +* Math (+)	+*
Biernacki (1976)	3-6	Toledo, Ohio	479 (174 NG, 305 G) (2 schools)	6 months	Schools matched on SES, race, and similar achievement in grade equivalents in reading and math for students in Grade 6. Students randomly selected from chosen schools. Pretest used to compute gain scores.	MAT		Rdg. +.13 Math +.20	+ .17

Klaus (1981)	4-6	LaCrosse, Wisconsin	433 (219 NG, 214 G)	3 yrs.	Pretest used to compute gain scores. IQ used as a covariate.	ITBS, SRAAS	Rdg. +.07 Math +.12 Lang. -.05	+ .05
Bradford (1972)	1-3	Detroit, Michigan	394 (299 NG, 93 G) (2 schools)	1 yr.	Students matched on sex, SES, and reading and math achievement. Pretest used to compute gain scores. IQ used as a covariate.	MAT	Rdg. + Math +*	+ *
Burtley (1974)	2-3	Woodberry, Illinois	302 (167 NG, 135 G)	2 yrs.	Schools matched on SES, ethnicity, size, and enrollment. Pretest used to compute gain scores.	MAT	Rdg. +.40* Math +.55*	+ .48*
Flowers (1977)	3	Westminster, Colorado	221 (99 NG, 122 G)	3 yrs.	Best school matches available among the remaining schools within the district based on SES. Students classified by SES.	SAT	Rdg. -.25 Lang. -.25	-.25
Kuhlman (1985)	2, 4, 6	Kansas	200 (50 OS, 50 IGE, 100 G)	2 yrs.	Students randomly selected from chosen schools. SES and number of parents used as covariates.	KCT	Rdg. 0 Math 0	0
Soumokol (1977)	3, 5	Columbia, Missouri	102 (2 schools)	2 yrs. (Gr. 3) 3 yrs. (Gr. 5)	Pretest and IQ used as covariates.	ITBS	Rdg. +.79** Math +.80**	+ .80**
Henn (1974)	4	Ohio	(24 schools) (Enrollment: 7,072 NG, 6,958 G)	2 yrs.	Schools matched on SES, available resources, and teachers' qualifications.	MAT, OST	Rdg. +.05 Math +.01	+ .03

Note. A key to the abbreviations used is provided as an appendix to this article.

\* $p < .05$

\*\* $p < .01$

with a wider range of instructional possibilities for their specific needs: small groups, one-to-one tutoring, or independent work. This finding supports the argument that selective use of individualized instruction can yield positive results for children's academic performance.

*Studies Lacking an Explicit Description of the Nongraded Program*

The last group of studies, summarized in Table 5, included six studies dated in the 1960s and six in the 1970s, all of which lacked an explicit description of the type of intervention applied to experimental schools and the characteristics of control schools. Two doubts confront any reader of these reports: To what extent was the nongraded label a good description of practices in the experimental situations, and what characteristics did control schools have that really made them fit a conventional description. The value of these studies was perhaps in putting to rest the idea that simply giving a school an innovative label, in this case *nongraded*, would have had some effect on student learning. These studies were included for the sake of completeness, but little can be learned from them.

In the largest matched equivalent study, Hickey (1962) found that students in nongraded primaries in seven Catholic schools learned significantly more after 3 years than did students in similar graded schools (total ES = +.46).

Among all other studies, only the smallest (Engel & Cooper, 1971) reported positive significant differences (total ES = +1.10). This study also differed from all others in its assurance that the schools compared belonged to extremes in an index of nongradedness.

Two medium-sized matched studies that lacked evidence of initial equality reported significant differences in favor of graded schools. Vogel and Bowers (1972) conducted a 1-year study of students in kindergarten through sixth grade. Pupils in graded schools scored higher on the standardized test (SAT) but significantly lower on measures of conceptual maturity (according to the Harris Draw-A-Person Test).

The second study that found significantly higher achievement in graded schools was also one of the lowest in methodological quality. Carbone (1961) compared the achievement of students in traditional graded schools to that in schools mentioned by Goodlad and Anderson (1959) as nongraded, controlling for IQ scores. The students involved were in grades four, five, and six—which is to say, 1, 2, or 3 years (respectively) after their experience in the nongraded primary. Further, there were substantial IQ differences between the two sets of students, and teacher questionnaires indicated very few differences between the two sets of teachers in reported classroom practices.

*Summary and discussion: Programs not adequately described.* This was the only group of research studies in which a trend was not evident. The median effect size was near zero (ES = +.02), but two studies found significantly positive effects, and two found significantly negative ones. The most serious limitation of these studies was the lack of descriptions that could have helped to interpret the findings. A closer look at the four studies that presented significant differences made the argument in favor of the nongraded programs more convincing. Both positive studies had greater methodological quality: Both had evidence of initial equality (students were matched on IQ), and Engel and Cooper (1971) even tested the validity of schools' labels. In contrast, each of the negative studies had some serious problems: inconsistency of findings or flaws in the experimental setting.

*Interactions With Study Features*

In addition to differentiating results according to categories of programs, assessments were made of the interaction of nongraded versus graded organization with other features of the studies. One finding, which buttresses the main conclusions of this review, was that program effects for the Joplin-like and comprehensive models were particularly strong and consistent in the higher quality studies—that is, the randomized and matched equivalent experiments. However, there were no consistent patterns with respect to effects at different grade levels (1–3 vs. 4–6) or subjects (reading, math, or language arts). Longer implementations (more than 1 year) were only inconsistently associated with larger effects.

Program effects declined according to the year of publication of the studies, but this is of course confounded with program types; the Joplin-like and comprehensive models without individualization were mostly studies reported in the 1960s, while most studies of nongraded programs emphasizing individualization, including studies of IGE, were reported in the 1970s.

**Does Nongrading Accelerate or Decelerate Student Progress?**

One of the principal rationales for nongrading is that it allows students to spend more time in the grades involved, if necessary, until they can reach a high level of performance or to spend less time if they are able to go more quickly than other students. Surprisingly, only one study actually assessed the degree to which nongraded students took nonnormative amounts of time to complete the primary or elementary grades. This was a study by McLoughlin (1970), which compared students in graded and nongraded primary programs in eight New York State school districts. The nongraded programs used flexible cross-class and cross-age grouping, teaching to homogeneous groups, and continuous-progress curricula; they would therefore probably fall into the comprehensive category defined earlier. The comparisons were made in 1964–1965 and again in 1965–1966. In 1964–1965, an average of 4.4% of students took an extra year to complete the primary grades in the nongraded schools; 4.6% took an extra year in the graded ones. In 1965–1966, 2.9% of nongraded students took an extra year, while 7.3% of graded students were retained. No students were accelerated in either type of program in 1964–1965, and  $\frac{1}{10}$  of one percent were accelerated in the nongraded schools in 1965–1966.

Put another way, 95.6% of the nongraded students made normative progress through the primary grades in 1964–1965, and 97.0% in 1965–1966. What this means is that, at least in the time and places studied by McLoughlin (1970), nongrading was not being used as a means of altering the amount of time students spent in the primary grades. On the contrary, in 1965–1966 slightly more students were decelerated by retention in the graded schools than the number that took the opportunity to spend more time in the primary grades offered by the nongraded structure.

McLoughlin (1970) also checked to see whether schools that had been implementing the nongraded plan over a longer period had more students who made nonnormative progress than newer nongraded programs. There was a slight (nonsignificant) trend, but it was in the opposite direction. First-year nongraded programs had somewhat more students making nonnormative progress than did schools that had implemented nongrading for 2–7 years.

If McLoughlin's findings (1970) apply to other implementations of the nongraded concept, this result has important methodological and substantive consequences.



**TABLE 5**  
*Studies lacking an explicit description of the nongraded program*

Article	Grades	Location	Sample size	Duration of program	Design	Test	Effect sizes		Total
							By achievement	By subject	
<i>Matched studies lacking evidence of initial equality</i>									
Hickey (1962)	3	Diocese of Pittsburgh	1348 (745 NG, 603 G) (14 schools)	3 yrs.	Schools matched on SES. Students matched on IQ.	MAT	Hi. +.31* Av. +.18 Lo. -.01	Rdg. +.24* Math +.68**	+ .46**
Lair (1975)	3	Richardson, Texas	463 (183 NG, 280 G) (12 schools)	3 yrs.	Random selection of 6 G and 2 NG schools. Students matched on readiness for learning scores.	CLTBS	Su. -.12 Hi. -.21 Av. -.36	Lang. -.09	-.09
Aigner (1961)	4	Bellevue, Washington	428 (214 NG, 214 G)	3 yrs.	Groups equated with the School and College Abilities Total test.	STEP	Hi. 0 Av. 0 Lo. 0	Rdg. (-) Math (-)	(-)
Mycock (1966)	1	Manchester, England	108 (4 schools)	1 yr.	Schools matched on size, resources, and staff ratio and quality. Students matched on age, sex, and intelligence.	NARA, PT		Rdg. 0 Math 0	0
Reid (1973)	4	Alabama	100 (50 NG, 50 G)	3 yrs.	Students matched on age, sex, and mental ability.	SAT	Hi. -.11 Av. -.01 Lo. -.05	Rdg. +.01 Math -.05 Lang. +.01	-.01

Williams (1966)	3	Hammond, Indiana	76 (38 NG, 38 G)	3 yrs.	Students matched on age, sex, and IQ.	SAT	Hi. +1.29* Lo. - 1.30*	Rdg. -.46* Math -.23	-.34
Engel & Cooper (1971)	6	Darmouth, Nova Scotia	40 (20 NG, 20 G) (2 schools)	6 yrs.	Schools selected according to an index for nongradedness. Students matched on IQ.	CAT		Rdg. + 1.20** Lang. + 1.02**	+1.10**
Herrington (1973)	6	Dade County, Florida	951 (16 Schools)	1 yr.	Schools randomly selected from SES-ranked lists. Classes randomly selected. Pretest used as covariate.	SAT		Rdg. (+) Math +	(+)
Vogel & Bowers (1972)	K-6	Evanston, Illinois	473 (224 NG, 249 G)	1 yr.	Teachers matched on sex, training, experience, and age level taught. Pretest used as covariate.	SAT		Composite - **	- **
Hopkins, Oldridge, & Williamson (1965)	3-4	Los Angeles County, California	330 (139 NG, 191 G) (4 schools)	3 yrs.	IQ used as covariate.	CRT		Rdg. + .02	+ .02
Carbone (1961)	4-6		244 (122 NG, 122 G) (6 schools)	At least 3 yrs.	Schools matched on SES. Classes randomly selected. Students matched on sex and age. IQ used as covariate.	ITBS		Rdg. - ** Math - ** Lang. - **	- **
Remacle (1970)	5-6	Brookings, South Dakota	128 (64 NG, 64 G) (1 school)	2 yrs. (Gr. 5) 1 yr. (Gr. 6)	Random selection of students in control group. IQ used as covariate.	ITBS		Rdg. + .24 GE Math + .37 GE* Lang. + .33 GE	+ .31 GE

*Note.* A key to the abbreviations used is provided as an appendix to this article.

\* $p < .05$

\*\* $p < .01$

Methodologically, there is a concern in studies of nongrading that, if nongraded students take more time to complete the primary grades, their test scores will be artificially increased. That is, if “third-year” students in a nongraded school were older on average than third graders in a graded school, this could explain any test-score advantage of the nongraded programs. It would be important to know this is not the case.

Substantively, McLoughlin’s findings (1970) may be seen as questioning one assumption of many advocates of nongrading, who often paint a picture of the low-achieving child proceeding happily and successfully through the grades, never particularly aware that he or she is taking 4 years to accomplish what his or her classmates are completing in 3 years. Yet students (and, more particularly, their parents) can count, and they know who their classmates were when they entered school. The pressures to have students make normative progress may be as strong in nongraded programs as in graded programs.

Yet the main thrust of Goodlad and Anderson’s rationale (1959) is not affected by McLoughlin’s findings (1970). It is still plausible that deceleration in a continuous-progress curriculum is preferable to retention. Further, in a flexible nongraded program it may be that students who would otherwise fall behind can be identified and given extra assistance so that they may catch up with their peers. The nongraded plan might be seen not as a way to give low achievers more time but rather as a way to use time and other resources more flexibly. A student who is not reading at the end of first grade might well be reading at the end of second grade if he or she receives extra help (and does not suffer the humiliation of repeating first grade).

### **Discussion**

As the nongraded elementary plan reappears in schools of the 1990s, it is important to learn about the history of this movement of 30 years ago. Most important, one needs to understand the achievement effects of nongraded organization and to understand the conditions under which achievement was or was not enhanced by this innovation.

A review of research on the nongraded elementary school is particularly needed today because there was little consensus on its effects in its own time. Only two reviewers examined portions of the literature, and they came to opposite conclusions. Pavan (1973, 1977, in press) concluded that the evidence favored the nongraded primary, while McLoughlin (1967) stated that most research showed no differences between graded and nongraded plans. Slavin and I conclude that, when their review methods are applied to a much larger set of studies, the evidence could be interpreted as confirming both Pavan’s and McLoughlin’s conclusions, contradictory though they are.

Table 6 summarizes the outcomes of the 57 studies that met the inclusion standards. Looking only at the box score of significant and nonsignificant positive and negative findings, one can read the results as supporting either McLoughlin’s (1967) negative conclusion or Pavan’s (1973, 1977, in press) positive one. McLoughlin argued that, because nonsignificant findings outnumbered significant positive ones, the effects of the nongraded primary were equivocal. Twenty-five years later, the proportions of significantly positive findings are like those he reported; only 20 of the 57 studies were significantly positive. Pavan came to the opposite conclusion in her reviews, noting that significant positive findings far outnumbered significant negative

**TABLE 6**

*Summary of effects by type of nongraded plan*

Type of program	Total studies	Significant positive	Nonsignificant positive	No difference	Nonsignificant negative	Significant negative	Median effect size
Joplin-Like	9	4	2	1	1	1	+ .46 (7)
Comprehensive	14	8	2	1	3	0	+ .34 (9)
Individualized	12	2	6	2	2	0	+ .02 (9)
IGE	10	4	3	2	1	0	+ .11 (6)
Unspecified	12	2	3	1	4	2	+ .01 (6)

*Note.* Number of studies in which an effect size could be computed is presented in parentheses.

ones. This is also true in the present review; only 3 studies significantly favored graded programs, while 20 favored nongraded ones.

However, the conclusions of the present review, which uses a best evidence synthesis, conform to neither McLoughlin's nor Pavan's conclusions. Instead, the evidence presented here supports a conclusion that the effects of nongraded programs depend on the type of program being implemented. Using median effect sizes rather than box scores, one sees that the positive effects of nongraded organization are most consistent and strongest when the program focuses on the vertical organization of the school and when nongrading is used as a grouping strategy but not as a framework for individualized instruction.

Four categories of nongraded programs were examined, in addition to one group of studies in which the nature of the nongraded program could not be determined. Studies in two of these categories clearly supported the nongraded plans. These are the Joplin-like programs, in which students are grouped across age lines in just one subject (usually reading), and the comprehensive programs, which involve cross-age grouping in many subjects but still rely on teacher-directed instruction. The median effect sizes for studies in these categories were clearly positive (+ .46 for Joplin-like programs, + .34 for comprehensive), and the best designed evaluations were the ones most likely to show the positive effects.

In contrast, nongraded programs that incorporated a great deal of individualized instruction (and correspondingly less teacher-directed instruction), including Individually Guided Education (IGE), were less consistently associated with achievement gain. This is not to say that these approaches reduce student achievement; rather, their effects are very inconsistent, generally neither helping nor hurting student achievement, with more studies finding positive than negative effects (especially in the case of IGE). Poorly described nongraded programs also had median effects very near zero, perhaps because experimental and control groups may not have differed in anything essential except label.

What accounts for the relatively consistent positive effects of the Joplin-like and comprehensive nongraded plans and the less consistent effects of programs incorporating individualization? At this remove of time from the flowering of the nongraded ideal, one can only speculate, but there are many more recent developments in educational research that suggest some possibilities.

The most obvious reason that incorporating a great deal of individualization might have reduced the effectiveness of the nongraded elementary school is suggested by research on individualized instruction itself, which has generally failed to support this innovation (e.g., see Bangert, Kulik, & Kulik, 1983; Horak, 1981; Miller, 1976; Rothrock, 1982). Correlational evidence from process-product studies of more and less effective teachers has consistently found that student learning is enhanced by direct instruction from teachers, as contrasted with extensive reliance on individualization, seatwork, and written materials (see Brophy & Good, 1986). Further, to the degree that the nongraded elementary school came to resemble the open school, the research finding few achievement benefits to this approach (e.g., Giaconia & Hedges, 1982) takes on increased relevance.

In its simplest forms, the nongraded elementary school has many likely benefits. By grouping students across age lines, it may allow teachers to reduce the number of within-class reading and math groups they teach at any given time, thereby reducing the need for independent seatwork and follow-up. In fact, in several of the evalua-

tions of Joplin-like programs, it was noted that cross-age groupings made within-class groupings (i.e., reading groups) unnecessary, so teachers could spend the class period teaching the entire class, with no need for seatwork unless they saw a specific need for it.

Another factor in the success of simple nongraded plans is the likelihood that flexible cross-age grouping allows teachers to fully accommodate instruction to the needs of each child in a particular subject while still delivering instruction to groups. Goodlad and Anderson's (1959, 1963) criticism of traditional ability grouping is that it does not truly reduce heterogeneity in the specific skill being taught. Grouping students within classes or within grades (in all but the largest elementary schools) does not provide enough opportunity to have group instruction closely tailored to student needs. Flexible cross-age grouping does provide such an opportunity, so the instructional costs of grouping (in terms of disruption, movement, and stigma for children in low groups) can perhaps be outweighed by the greater opportunity to adapt instruction to the precise needs of students and to continue to adapt to students' needs by examining and changing groupings at frequent intervals (see Slavin, 1987).

If the effectiveness of nongraded organization is due to increased direct instruction delivered at students' precise instructional level, then it is easy to see how a move to greater individualization would undermine these effects. Individualized instruction, learning stations, learning activity packets, and other individualized or small group activities reduce direct instruction time with little corresponding increase in appropriateness of instruction to individual needs (in comparison to the simpler cross-age grouping plans).

It is difficult to assess the impact of one of the key rationales given for the nongraded plan throughout its history, the opportunity to allow at-risk students to take as much time as they need to complete the primary or elementary grades without the use of retention. An early study by McLoughlin (1970) found that self-described nongraded programs did not generally take advantage of the opportunity to let low achievers take more time, but one does not know if McLoughlin's findings would apply to most nongraded programs implemented now or in the past. Clearly, however, the effectiveness of the simpler nongraded programs does not depend on the opportunity to accelerate or decelerate student progress, since most studies found positive effects in the first year of implementation, before any acceleration or deceleration could take place.

This discussion is, as noted earlier, completely speculative. There is much more we would have liked to know about how nongraded programs were actually implemented in the 1950s, 1960s, and 1970s. The return of the nongraded idea in the 1990s may, however, answer many questions. But assessments of current forms of nongrading as well as component analyses are necessary to understand which elements of nongrading account for the program's effects, and studies combining qualitative and quantitative methods are necessary to understand both what really changes in nongraded schools and what differences these changes make in student achievement.

*Is Earlier Research on the Nongraded Elementary School Relevant Today?*

How relevant is research on the nongraded elementary school to education today? Many of the problems that the nongraded elementary school was designed to solve still exist, and the reemergence of nongraded programs appears to be due in large

part to concern about these problems, especially the tension between retention and social promotion and rejection of traditional forms of ability grouping. Yet there are also many differences between education today and that of 30 years ago. The general perception that both individualized instruction (e.g., Bangert et al., 1983; Horak, 1981) and the open classroom (e.g., Giaconia & Hedges, 1982) failed in their attempt to increase student achievement means that it is unlikely that the nongraded elementary schools of the 1990s will, like those of the early 1970s, embrace these methods. As a result, it is more likely that the nongraded programs of the 1990s will resemble the simpler forms found in this review to be instructionally effective. Yet there are other developments in North American education today that will certainly influence the forms taken by the nongraded programs, their effects on achievement, and their ultimate impact on educational practice. The movement toward developmentally appropriate early childhood education and its association with nongrading means that the nongraded primary school of the 1990s will often incorporate 4- and 5-year-olds (earlier forms rarely did so) and that instruction in nongraded primary programs will probably be more integrated and thematic, and less academically structured or hierarchical, than other schools. A proposal for nongraded primary programs of this type was recently made by Katz et al. (1991). In other words, like in the early 1970s, the effectiveness of the nongraded school organization plan may become confounded with innovative instructional methods. Whether these instructional methods will have positive or negative effects on ultimate achievement is currently unknown.

The ultimate impact of the nongraded ideal will also have much to do with rapidly unfolding changes in assessment and accountability. One reason for the increase in retention, prefirst, and other extra-year programs in the late 1970s and early 1980s was greatly increased accountability pressures in U.S. schools. Retaining more students has a strong (though short-lived), positive impact on achievement test scores reported by grade (not age), because the children taking the tests are older (see Allington & McGill-Franzen, 1992; Slavin & Madden, 1991). There is currently widespread concern about high retention rates (Shepard & Smith, 1989), yet returning to social promotion would greatly reduce test scores in districts currently retaining many students. If the nongraded elementary school emerges as a means of giving low achievers more time in the elementary grades, it may be favored by the current policies of reporting test scores by grade (for the same reasons that they favor retention). On the other hand, if high-stakes accountability systems begin to report achievement by age (e.g., as does the National Assessment of Educational Progress), this advantage may not become a factor.

Clearly, there is a need for much more research on the nongraded elementary school as it is being implemented today. Because of scientific conventions of the time, most of the earlier research reviewed here was strong in experimental design (most studies used random assignment or careful matching of experimental and control groups) but weak in description of the independent variable—that is, the characteristics of the nongraded and graded schools. Research done today must be strong in both dimensions. Goodlad and Anderson (1963) emphasized that the nongraded plan addresses only vertical organization, not instruction. Yet, as this review has shown, differences in instructional methods between nongraded and graded schools may account for differences (or nondifferences) in outcomes.

Research on the nongraded elementary school offers a fascinating glimpse into the history and ultimate fate of a compelling innovation. The return of this idea after

nearly 20 years of dormancy is fascinating as well. This review concludes that the evidence from the first cycle of research on the nongraded elementary school supports use of simpler forms of the model and certainly supports the need and potential fruitfulness of further experimentation. Yet there is a cautionary note in this review as well. Good ideas can be undermined by complexification over time. A constant cycle of experimentation, research, evaluation, revision, and continued experimentation is necessary to build compelling ideas into comprehensive, effective plans for school organization and instruction.

## APPENDIX

### *Abbreviations used in the tables*

---

+	Results clearly favor nongraded programs
(+)	Results generally favor nongraded programs
0	No trend in results
(-)	Results generally favor graded programs
-	Results clearly favor graded programs
AG	Ability Grouping
CAT	California Achievement Test
CRT	California Reading Test
CTBS	Canadian Test of Basic Skills
CLTBS	California Test of Basic Skills
DLRT	Durrell Listening-Reading Test
GPRT	Gates Primary Reading Test
IGE	Individually Guided Education
IGI	Individually Guided Instruction
ITBS	Iowa Test of Basic Skills
G	Graded Program
GE	Grade Equivalent
KCT	Kansas Competency Test
LCRT	Lee Clark Reading Test
MAT	Metropolitan Achievement Test
MRT	Metropolitan Reading Test
NARA	Neale Analysis for Reading Ability
NG	Nongraded Program
OS	Open-Area School
OST	Ohio Survey Test
PT	Piaget-Type Test
Rdg.	Reading
SAT	Stanford Achievement Test
SRAAS	Science Research Associates Achievement Series
STEP	Sequential Test of Educational Programs

---

## References

- Aigner, B. W. (1961). *A statistical analysis of achievement differences of children in a nongraded primary program and traditional classrooms*. Unpublished doctoral dissertation, Colorado State University, Fort Collins.
- Allington, R. L., & McGill-Franzen, A. (1992). Does high-stakes testing improve school effectiveness? *ERS Spectrum*, 10 (2), 3-12.
- Bangert, R., Kulik, J., & Kulik, C. (1983). Individualized systems of instruction in secondary schools. *Review of Educational Research*, 53, 143-158.



- Barr, R. (1990). The social organization of literacy instruction. In S. McCormick & J. Zutell (Eds.), *Thirty-ninth yearbook of the National Reading Conference* (pp. 19–33). Chicago: National Reading Conference.
- Biernacki, G. J. (1976). *An evaluation of an IGE/MUS-E inner-city elementary school based on selected criteria*. Unpublished doctoral dissertation, University of Toledo.
- Bockrath, S. M. B. (1958). *An evaluation of the ungraded primary as an organizational device for improving learning in Saint Louis Archdiocesan Schools*. Unpublished doctoral dissertation, St. Louis University.
- Bowman, B. L. (1971). *A comparison of pupil achievement and attitude in a graded school with pupil achievement and attitude in a nongraded school 1968–69*. Unpublished doctoral dissertation, University of North Carolina, Chapel Hill.
- Bradford, E. F. (1972). *A comparison of two methods of teaching in the elementary school as related to achievement in reading, mathematics, and self-concept of children*. Unpublished doctoral dissertation, Michigan State University, East Lansing.
- Brody, E. B. (1970). Achievement of first- and second-year pupils in graded and nongraded classrooms. *Elementary School Journal*, 70, 391–394.
- Brophy, J. E., & Good, T. L. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 328–375). New York: Macmillan.
- Buffie, E. G. W. (1962). *A comparison of mental health and academic achievement: The nongraded school vs. the graded school*. Unpublished doctoral dissertation, Indiana University.
- Burchyett, J. A. (1972). *A comparison of the effects of non-graded, multi-age, team-teaching vs. the modified self-contained classroom at the elementary school level*. Unpublished doctoral dissertation, Michigan State University, East Lansing.
- Burtley, N. (1974). *A comparison of teacher characteristics and student achievement in individually guided education (IGE) and traditional inner city elementary schools*. Unpublished doctoral dissertation, Michigan State University, East Lansing.
- Carbone, R. F. (1961). A comparison of graded and non-graded elementary schools. *Elementary School Journal*, 61, 82–88.
- Case, D. A. (1970). *A comparative study of fifth graders in a new middle school with fifth graders in elementary self-contained classrooms*. Unpublished doctoral dissertation, University of Florida, Gainesville.
- Chalfant, L. S. (1972). *A three year comparative study between students in a graded and nongraded secondary school*. Unpublished doctoral dissertation, Utah State University, Logan.
- Chastain, C. S. (1961). *An experimental study of the gains in achievement in arithmetic and reading made by the pupils in the intermediate grades in the Rangely, Colorado, Elementary School platoons who were instructed in traditional classrooms, in achievement and in non-graded classrooms*. Unpublished doctoral dissertation, University of Northern Colorado, Greeley.
- Enevoldsen, C. L. (1961). *An evaluation of the ungraded primary program in selected schools in the Lincoln, Nebraska, Public School system*. Unpublished doctoral dissertation, University of Nebraska, Teachers College, Lincoln.
- Engel, B. M., & Cooper, M. (1971). Academic achievement and nongradedness. *The Journal of Experimental Education*, 40, 24–26.
- Flowers, J. R. (1977). *A comparative study of students in open space individually guided education (IGE) and traditional schools*. Unpublished doctoral dissertation, University of Northern Colorado, Greeley.
- Floyd, C. (1954). Meeting children's reading needs in the middle grades: A preliminary report. *Elementary School Journal*, 55, 99–103.
- Giaconia, R. M., & Hedges, L. V. (1982). Identifying features of open education. *Review of Educational Research*, 52, 579–602.

- Givens, H., Jr. (1972). *A comparative study of achievement and attitudinal characteristics of black and white intermediate pupils in individualized, multigrade and self-contained instructional programs*. Unpublished doctoral dissertation, St. Louis University.
- Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills: Sage.
- Goodlad, J. I., & Anderson, R. H. (1959). *The nongraded elementary school*. New York: Harcourt, Brace & Co.
- Goodlad, J. I., & Anderson, R. H. (1963). *The nongraded elementary school* (rev. ed.). New York: Harcourt, Brace, & World.
- Goodlad, J. I., & Anderson, R. H. (1987). *The nongraded elementary school*. New York: Teachers College Press. (Original work published 1963)
- Guarino, A. R. (1982). *An investigation of achievement, self-concept, and school related anxiety in graded and nongraded elementary schools*. Unpublished doctoral dissertation, State University of New Jersey, Rutgers.
- Gumpper, D. C., Meyer, J. H., & Kaufman, J. J. (1971). *Nongraded elementary education: individualized learning-teacher leadership-student responsibility*. University Park: Pennsylvania State University, Institute for Research on Human Resources. (ERIC Document Reproduction Service No. ED 057 440)
- Halliwell, J. W. (1963). A comparison of pupil achievement in graded and nongraded primary classrooms. *The Journal of Experimental Education*, 32, 59–64.
- Hart, R. H. (1962). The nongraded primary school and arithmetic. *The Arithmetic Teacher*, 9, 130–133.
- Heathers, G. (1967). *Organizing schools through the dual progress plan*. Danville, IL: Interstate.
- Henn, D. C. (1974). *A comparative analysis of language arts and mathematics achievement in selected IGE/MUS-E and non-IGE/MUS-E programs in Ohio*. Unpublished doctoral dissertation, University of Cincinnati.
- Herrington, A. F. (1973). *Perceived attitudes and academic achievement of reported graded and nongraded sixth-year students*. Unpublished doctoral dissertation, University of Miami, Coral Gables.
- Hickey, S. M. P. (1962). *An analysis and evaluation of the ungraded primary program in the diocese of Pittsburgh*. Unpublished doctoral dissertation, Fordham University, Bronx.
- Higgins, J. J. (1980). *A comparative study between the reading achievement levels of students in a combination/ungraded class and students in a graded class*. Unpublished doctoral dissertation, George Peabody College for Teachers of Vanderbilt University.
- Hopkins, K. D., Oldridge, O. A., & Williamson, M. L. (1965). An empirical comparison of pupil achievement and other variables in graded and ungraded classes. *American Educational Research Journal*, 2, 207–215.
- Horak, V. M. (1981). A meta-analysis of research findings on individualized instruction in mathematics. *Journal of Educational Research*, 74, 249–253.
- Ingram, V. (1960). Flint evaluates its primary cycle. *Elementary School Journal*, 61, 76–80.
- Jacquette, F. C. (1959). *A five year study to determine the effects of the ungraded classroom organization on reading achievement in Grand Junction, Colorado*. Unpublished field study, University of Northern Colorado, Greeley.
- Jeffreys, J. S. (1970). *An investigation of the effects of innovative educational practices on pupil-centeredness of observed behaviors and on learner outcome variables*. Unpublished doctoral dissertation, University of Maryland, College Park.
- Jones, J. C., Moore, J. W., & Van Devender, F. (1967). A comparison of pupil achievement after one and one-half and three years in a nongraded program. *The Journal of Educational Research*, 61, 75–77.
- Karweit, N. L., & Wasik, B. A. (in press). Extra-year kindergarten programs and transitional first grades. In R. E. Slavin, N. L. Karweit, & B. A. Wasik (Eds.), *Preventing early school failure: Research on effective strategies*. Boston: Allyn & Bacon.

- Katz, L. G., Evangelou, D., & Hartman, J. A. (1991). *The case for mixed-aged grouping in early childhood education*. Washington, DC: National Association for the Education of Young Children.
- Kierstead, R. (1963). A comparison and evaluation of two methods of organization for the teaching of reading. *The Journal of Educational Research*, 56, 317–321.
- Killough, C. K. (1971). *An analysis of the longitudinal effects that a nongraded elementary program, conducted in an open-space school, had on the cognitive achievement of pupils*. Unpublished doctoral dissertation, University of Houston.
- Klaus, W. D. (1981). *A comparison of student achievement in individually guided education programs and non-individually guided elementary school programs*. Unpublished doctoral dissertation, University of Missouri, Columbia.
- Klausmeier, H. J., Quilling, M. R., Sorenson, J. S., Way, R. S., & Glasrud, G.R. (1971). *Individually guided education and the multiunit elementary school: Guidelines for implementation*. Madison: University of Wisconsin, Wisconsin Research & Development Center for Cognitive Learning.
- Klausmeier, H. J., Rossmiller, R. A., Saily, M. (Eds.). (1977). *Individually Guided Elementary Education: Concepts and Practices*. New York: Academic.
- Kuhlman, C. L. (1985). *A study of the relationship between organizational characteristics of elementary schools, student characteristics, and elementary competency test results*. Unpublished doctoral dissertation, University of Kansas, Lawrence.
- Lair, D. P. (1975). *The effects of graded and nongraded schools on student growth*. Unpublished doctoral dissertation, Baylor University, Waco.
- Lawson, R. E. (1973). *A comparison of the development of self-concept and achievement in reading of students in the first, third, and fifth year of attendance*. Unpublished doctoral dissertation, Ball State University, Muncie, IN.
- Levine, D. U., & Eubanks, E. E. (1986–1987). Achievement improvement and non-improvement at concentrated poverty schools in big cities. *Metropolitan Education*, 3, 92–107.
- Machiele, R. B. (1965). A preliminary evaluation of the non-graded primary at Leal School, Urbana. *Illinois School Research*, 1, 20–24.
- Maresh, R. T. (1971). *An analysis of the effects of vertical grade groups on reading achievement and attitudes in elementary schools*. Unpublished doctoral dissertation, University of North Dakota, Grand Forks.
- McLoughlin, W. P. (1967). *The nongraded school: A critical assessment*. New York: State Education Department.
- McLoughlin, W. P. (1970). Continuous pupil progress in the non-graded school: Hope or hoax? *Elementary School Journal*, 71 (2), 90–96.
- McPartland, J. M., Coldiron, J. R., & Braddock, J. H. (1987). *School structures and classroom practices in elementary, middle, and secondary schools* (Tech. Rep. No. 14). Baltimore, MD: Johns Hopkins University, Center for Research on Elementary and Middle Schools.
- Miller, R. L. (1976). Individualized instruction in mathematics: A review of research. *The Mathematics Teacher*, 69, 345–351.
- Moore, D. I. (1963). *Pupil achievement and grouping practices in graded and ungraded primary schools*. Unpublished doctoral dissertation, University of Michigan, East Lansing.
- Morris, V. R. (1968). *An evaluation of pupil achievement in a nongraded primary plan after three, and also five years of instruction*. Unpublished doctoral dissertation, Lehigh University, Bethlehem, PA.
- Muck, R. E. S. (1966). *The effect of classroom organization on academic achievement in graded and nongraded classes*. Unpublished doctoral dissertation, State University of New York, Buffalo.
- Mycok, M. A. (1966). *A comparison of vertical grouping and horizontal grouping in the infant school*. Unpublished master's thesis, University of Manchester.
- National Association for the Education of Young Children. (1989). *Appropriate education in the primary grades*. Washington, DC: Author.

- Otto, H. J. (1969). *Nongradedness: An elementary school evaluation*. Austin: University of Texas.
- Pavan, B. N. (1972). *Moving elementary schools toward nongradedness: Commitment, assessment, and tactics*. Unpublished doctoral dissertation, Harvard University, Graduate School of Education, Cambridge.
- Pavan, B. N. (1973). Good news: Research on the nongraded elementary school. *Elementary School Journal*, 73 (6), 333–342.
- Pavan, B. N. (1977). The nongraded elementary school: Research on academic and mental health. *Texas Tech Journal of Education*, 4, 91–107.
- Pavan, B. N. (in press). Recent research on nongraded schools: Benefits for achievement, mental health, and disadvantaged students. *Educational Leadership*.
- Perrin, J. D. (1969). *A statistical and time change analysis of achievement differences of children in a nongraded and a graded program in selected schools in the Little Rock Public Schools*. Unpublished doctoral dissertation, University of Arkansas, Little Rock.
- Price, D. A. (1977). *The effects of individually guided education (IGE) processes on achievement and attitudes of elementary school students*. Unpublished doctoral dissertation, University of Missouri, Columbia.
- Ramayya, D. P. (1972). *Achievement skills, personality variables, and classroom climate in graded and nongraded elementary schools*. Nova Scotia: Dartmouth Public Schools.
- Reichardt, C. S. (1979). The statistical analysis of data from nonequivalent group designs. In T. C. Cook & D. T. Campbell (Eds.), *Quasi-experimentation: Design and analysis issues for field settings* (pp. 147–205). Chicago: Rand McNally.
- Reid, B. C. (1973). *A comparative analysis of a nongraded and graded primary program*. Unpublished doctoral dissertation, University of Alabama, Tuscaloosa.
- Remacle, L. F. (1970). *A comparative study of the differences in attitudes, self-concept, and achievement of children in graded and nongraded elementary schools*. Unpublished doctoral dissertation, University of South Dakota, Vermillion.
- Ross, G. A. (1967). *A comparative study of pupil progress in ungraded and graded primary programs*. Unpublished doctoral dissertation, Indiana University.
- Rothrock, D. (1982). The rise and decline of individualized instruction. *Educational Leadership*, 39, 528–531.
- Schneiderhan, R. M. (1973). *A comparison of an individually guided education (IGE) program, an individually guided instruction (IGI) program, and a traditional elementary educational program at the intermediate level*. Unpublished doctoral dissertation, University of Minnesota, Duluth.
- Shepard, L. A., & Smith, M. L. (Eds.). (1989). *Flunking grades: Research and policies on retention*. New York: Falmer.
- Sie, M.S. (1969). *Pupil achievement in an experimental nongraded elementary school*. Unpublished doctoral dissertation, Iowa State University of Science and Technology, Ames.
- Skapski, M. K. (1960). Ungraded primary reading program: An objective evaluation. *Elementary School Journal*, 61, 41–45.
- Slavin, R. E. (1986). Best-evidence synthesis: An alternative to meta-analytic and traditional reviews. *Educational Researcher*, 15 (9), 5–11.
- Slavin, R. E. (1987). Ability grouping and student achievement in elementary schools: A best-evidence synthesis. *Review of Educational Research*, 57, 293–336.
- Slavin, R. E. (1990). Ability grouping and student achievement in secondary schools: A best-evidence synthesis. *Review of Educational Research*, 60, 471–499.
- Slavin, R. E., & Madden, N.A. (1991). Modifying Chapter 1 program improvement guidelines to reward appropriate practices. *Educational Evaluation and Policy Analysis*, 13, 369–379.
- Snake River School District. (1972). *Curriculum change through nongraded individualization*. Blackfoot, ID: School District 52.
- Soumokol, P. O. (1977). *Comparison of cognitive and affective dimensions of individually guided education (IGE) and standard elementary school programs*. Unpublished doctoral dissertation, University of Missouri, Columbia.

- Vogel, F. S., & Bowers, N. D. (1972). Pupil behavior in a multiage nongraded school. *The Journal of Experimental Education*, 41, 78-86.
- Walker, W. E. (1973). *Long-term effects of graded and nongraded primary programs*. Unpublished doctoral dissertation, George Peabody College for Teachers of Vanderbilt University.
- Ward, D. N. (1969). *An evaluation of a nongraded school program in grades one and two*. Unpublished doctoral dissertation, University of Texas, Austin.
- Williams, W. (1966). Academic achievement in a graded school and in a non-graded school. *Elementary School Journal*, 66, 135-139.
- Willis, S. (1991). Breaking down grade barriers: Interest in nongraded classrooms on the rise. *ASCD Update*, 33 (3), 1, 4.
- Wilt, H. J. (1970). *A comparison of student attitudes toward school, academic achievement, internal structures and procedures: The nongraded school vs. the graded school*. Unpublished doctoral dissertation, University of Missouri, Columbia.
- Zerby, J. R. (1960). *A comparison of academic achievement and social adjustment of primary school children in the graded and nongraded school programs*. Unpublished doctoral dissertation, Pennsylvania State University, University Park.

#### **Authors**

ROBERTO GUTIÉRREZ is PhD Candidate, Johns Hopkins University, Department of Sociology, Baltimore, MD 21218. He specializes in human development, national development, and school organization.

ROBERT E. SLAVIN is Co-Director, Elementary School Programs, Johns Hopkins University, Center for Research on Effective Schooling for Disadvantaged Students, 3505 N. Charles St., Baltimore, MD 21218. He specializes in cooperative learning, school organization, programs for students at risk, and research review.

Received April 14, 1992

Revision received July 2, 1992

Accepted July 29, 1992