A HISTORY OF TRACKING IN THE PUBLIC HIGH SCHOOL SCIENCE CLASSROOM OVER THE TWENTIETH CENTURY

by

SARA RAVEN

(Under the Direction of J. Steve Oliver)

ABSTRACT

A comprehensive history of tracking students by ability in the public high school science classroom is examined using peer reviewed journal articles and books published over the last century. In this qualitative historical study, the major educational reforms and events from the twentieth century were analyzed to determine their impact on ability grouping practices. Additionally, the study sought to determine the commonalities between current and past ability grouping practices. It was determined that, despite the many changes that have occurred in tracking procedures over the past century, may elements remain from ability grouping's conception that can inform current practices and help determine a future course of action.

INDEX WORDS: Academic tracking, Ability grouping, Secondary science education, Historical methodology

A HISTORY OF TRACKING IN THE PUBLIC HIGH SCHOOL SCIENCE CLASSROOM OVER THE TWENTIETH CENTURY

by

SARA RAVEN

B.S.Ed., The University of Georgia, 2008

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF ARTS

ATHENS, GEORGIA

2010

A HISTORY OF TRACKING IN THE PUBLIC HIGH SCHOOL SCIENCE CLASSROOM OVER THE TWENTIETH CENTURY

by

SARA RAVEN

Major Professor:	J. Steve Oliver
Committee:	David Jackson Norman Thomson

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2010

ACKNOWLEDGEMENTS

Thank you:

To my parents, who still really want me to be a doctor but will just have to wait three more years. To my sister, who will probably never write a thesis because I complained about mine so much. To my Grandma Lenore, who always had more faith in me to finish than I did in myself. To my Grandma Sheila, who never stopped trying to tell me that a B.S. meant one thing and an M.S. was just more of the same. To my co-workers, who helped me catch more typos then i cen cownt. To my advisor, who always appreciated my sense of humor and lack of self-esteem. To my students, who always knew the days I had to finish something important, and somehow had more questions on those days than on any others. And finally, to my wife, who loves me more than life itself; you know I feel the same.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Purpose

Since the beginning of the twentieth century, the effects of tracking by ability, or ability grouping, in high school science classrooms have been the subject of debate within academia and educational settings. Despite over a century of confusion, the answer to whether or not tracking is beneficial or harmful has yet to be agreed upon by many of the stakeholders. Current studies present opposing views while conflicting meta-analyses are regularly released. The reason for this lack of coherent understanding, however, may lie in the wide variation of ways in which ability grouping is practiced. Some schools claim to track students solely by intelligence, while others allow a modicum of student choice in the groups to which they are assigned. Some have formal ability grouping policies while some claim not to group students by ability at all, even though studies have shown that many schools which do not have formal ability grouping policies may practice grouping informally (Hoffer, T. B., 1992; Rees, Brewer, & Argys, 2000). These variations introduce uncertainty into the ability grouping debate and create difficulties in identifying the effects of ability grouping on student performance.

I analyzed present day ability grouping practices and several education reforms over the past century that have affected the way ability grouping is viewed and practiced. I began with the Committee of Ten's perspective on curriculum differentiation, followed by the *Cardinal Principles of Education*, the progressive education movement and compulsory education. Up

until the middle of the century, science classroom instruction was not an area on which ability grouping was focused: Sputnik changed all of that. The more the United States began to believe that America's students were falling behind those of the Soviet Union, the more voraciously gifted students in science were pursued to become science specialists. In addition to the Space Race, the second half of the century introduced many other reforms, including the civil rights movement and several important government-introduced doctrines.

Until forty years ago, the majority of academic papers expressed the opinion that ability grouping was in the best interest of all students. That opinion began to shift around the late 1960's because of several events, one of which was the desegregation of schools. Ability grouping practices were no longer viewed as the solution to excellence in education, but a method used to stratify students, both economically and socially. Studies began to show that tracking closely followed a caste system: the higher the parents' socioeconomic status, the more likely their child would attend the best schools and be placed in the highest ability groups. This trend shed light on the ability grouping dilemma. Although the purpose of ability grouping may have begun as a way to stratify students based on intelligence, it became apparent that students were not separated by this alone.

Looking back over the past century provides a perspective on ability grouping that is desperately needed in education. Without understanding tracking's past, it can be difficult to understand all facets of the problems plaguing it. The research for my thesis concentrated on formal ability grouping practices, as they relate to science teaching when applicable, over the past century. This involved how students were separated into groups and the differentiation each group received. The perceptions of these groups of students were also used. I focused

solely on ability grouping practices and the changes they have undergone since the turn of the twentieth century. Whether or not present-day practices have commonalities with past practices was examined. This document details how I sought to understand, analyze and interpret, through historical methodology, the evidence on this topic.

Subjectivity Statement

Some of my earliest memories of school are from the TAG (talented and gifted) program in my elementary school. Twice a week, our TAG teacher would pull me and the rest of the TAG students out from our regular schedule and we would spend a few hours on other activities. Sometimes it was word puzzles, or number games, or even making Rube Goldberg machines. After a few hours of TAG, we would be returned to our classrooms, where most of the students that had been left behind looked bored and tired, whereas we were excited to tell everyone about our time in the TAG room.

In middle school and high school, I didn't have TAG anymore, but gifted and honors classes. Regardless of which name was used, the reason was obvious. Students in the gifted classes were smarter and quicker. Everyone in the school knew that. And if you were not in the gifted classes, then you were just in regular biology or regular language arts. At the time, I just assumed that I was placed in these classes because of my intelligence. Somehow, I had been blessed with the smart gene, and this was my reward. This reward took me through high school and college, where I graduated with a science education degree. It wasn't until graduate school that I began to wonder, why me?

My first teaching job was this past year, in a small alternative school for students who were not as fortunate as I was during my childhood. These students were not, for the most

part, in gifted classes, or honors, or even regular. They were also not well-off economically; some were pregnant, had family problems, or were required to work two or three jobs to support their family. I was suddenly aware of two things: that I had been supremely lucky and that maybe it wasn't about what you know, but how you were raised.

For several months, I fell into a sort of depression. It was hard for me to believe that the environment I had been raised in was possibly false to a certain extent. An affluent, white student, taken care of in all of the basic necessities of life and then some, can be naïve, and I felt as though somebody had taken away the film through which I saw the world. What worried me the most, however, was that there was something wrong with our public education system. Why did affluent students receive a better education? Why were they enrolled in honors classes more often than students from poorer backgrounds? My background in science would not allow me to believe that it was just coincidence. I needed to know why minorities and students from lower socioeconomic backgrounds were not being given a fair chance.

When I first began teaching, I also enrolled in an MA program for science education. During my second semester, I took a course about sociological theories in education that allowed me to investigate the effects of tracking, ability grouping, and curriculum differentiation, and whether or not these methods of education were legitimate methods of schooling. I decided to center my thesis on this subject because, although there have been many studies pertaining to tracking, its legitimacy remains to be seen. It is obvious that I am biased to a certain extent and while I believe that tracking, ultimately, may not be the best solution for our schools, I was more than willing to discover the answer, regardless of my beliefs.

Rationale

Tracking/Ability Grouping

I chose to study ability grouping because it so deeply pervades the United States education system and is a practice that can commonly be found in public schools. Ability grouping, or tracking, is the process of grouping students by their perceived ability within schools. Ability, in most cases today, is a measure of standardized test scores, past performance in classes, and IQ scores. Ability grouping's conception is one that is acutely connected to the principles on which this country was founded. The pursuit of individual success, the American Dream, encapsulates the idea that anyone can achieve given enough drive and gusto. What this means in terms of education is that a child, no matter his or her socioeconomic status, can achieve great intellectual gains. The basic idea is, the more determined and motivated the student, the more likely he or she will be placed in the highest ability group. Is this an accurate depiction of ability grouping, however, or merely America's perception of it?

Over the past forty years, more attention has been paid to lower ability groups as a result of studies that linked socioeconomic status to ability group placement (Cheng, Brizendine, & Oakes, 1979; Gamoran, 1992a; Johnston, 1967). Many of these studies found that ability group placement has more to do with students' race, class or gender than with any other factor, such as intelligence and past school performance. Therefore, a minority student of lower socioeconomic status is more likely to be placed in the low ability group, regardless of his or her actual ability. Were these students placed in the lower group unfairly, through no fault of their own? Similarly, proponents of ability grouping have extolled its virtues, focusing

more on the gifted student. Is it inequitable to subject a student to a slower-paced classroom for the sake of equality, or does that do a disservice to the mind of a potential genius? These questions, though studied intensely, have yet to be answered. While I did not claim to be able to answer them, it is conceivable that by attempting to look back at the past, an answer to any one of these questions may be found.

Educational Reforms and Historical Research

The process of ability grouping has changed greatly over the past century. In that time, groups such as the Committee of Ten and the NEA Committee have been formed to evaluate the effect of ability grouping on students in the sciences and other subject areas; doctrines have been proposed to offer recommendations or condemn practices, and entire movements have been born that may have completely changed ability grouping practices into what can be seen today. Educational reforms are not just a lens through which our education history can be viewed, but a part of our present-day situation and, very likely, our future. Richard Marius, author of Writing about History, wrote that, despite the well-known adage, history does not repeat itself and that studying history does not help predict the future, nor does it offer broad insights into the present (Marius, 1989). Unfortunately, I had to disagree.

Public school education has a history that resembles the waves of the ocean, with initiatives and ideas disappearing and reappearing regularly. Bybee made note of this when he stated that parallels existed between education in the 1980's and the 1950's (Bybee, 1982). Labaree detailed the progressive education movement and showed that administrative progressives' victory in the mid-twentieth century still affects education today in very significant ways (Labaree D. F., 2005). This is not to say that everything stays the same.

America's schools are vastly changed from when they began, but the elements, the perspectives that shape lessons and discipline, administration and education, are still there. They float around schools, just waiting for someone to pull them out of thin air, as if the concept were entirely new. History is always around us, and only by studying its details, can we truly get a sense of where we've been, and where we can go.

Research Questions

The following were research questions that I attempted to answer over the course of this study. Although broad, they gave focus and direction to the research:

- What are the characteristics of academic tracking practices in present day public high school science classrooms?
- Do present day academic tracking practices have commonalities with practices that characterized classroom tracking earlier in the 20th century and, if so, what commonalities?

Goal

While the following question could not be answered as definitively as the research questions stated above, it helped to guide the study's conclusions and recommendations.

- How can the history of ability grouping practices in public high school science classrooms inform present day practices in schools?

Overview

This study was an historical investigation of the formal education practice known as ability grouping or tracking and an examination of how this practice has been changed over the past century. Ability grouping practices have been affected by many educational reforms,

several of which can still be found in present day public high schools. Ability grouping for the science classroom has been particularly shaped by these reforms. This thesis followed the timeline presented earlier to evaluate ability grouping practices. Utilizing my research, I assessed the characteristics of ability grouping as they were enacted a century ago and analyzed how those characteristics were represented in the characteristics of practices that were enacted in the intervening time. Additionally, I examined how ability grouping practices in science have been affected by educational reforms over the past century. Finally, I evaluated how past ability grouping practices can inform current ones.

CHAPTER 2

LITERATURE REVIEW

Definition of Ability Grouping

There are many different definitions of ability grouping accepted in academia, some broad and some narrow. Kulik and Kulik defined ability grouping as "the practice of organizing classrooms in graded schools to combine children who are similar in ability" (Kulik & Kulik, 1982, p. 415). Gamoran used ability grouping to refer to "divisions among students for particular subjects, such as special class assignments for math or within-class groups for reading" (Gamoran, 1992a, p. 11). Tracking, Gamoran purported, represents a different practice, which he defined as "broad, programmatic divisions that separate students for all academic subjects." Linchevski and Kutscher, however, identified tracking as a "method[s] of grouping students in the same grade…on a subject-by-subject basis" (Linchevski & Kutscher, 1998, p. 533). For the purposes of this study, I focused more on Slavin's definition of ability grouping:

> Ability grouping is defined as any school or classroom organization plan that is intended to reduce the heterogeneity of instructional groups; in between-class ability grouping the heterogeneity of each class for a given subject is reduced, and in within-class ability grouping the heterogeneity of groups within the class (e.g., reading groups) is reduced (Slavin R. E., 1990, p. 471).

While working with Slavin's definition allows a more encompassing view, it is because of broad and varied definitions that so many schools practice informal ability grouping, possibly widening the achievement gap unintentionally.

Whether ability grouping is practiced formally or informally leads to another complication in determining its range of practice. Rees, Brewer and Argys discussed issues that plague researchers trying to determine the effect size of ability grouping. One such issue is the concept of informal ability grouping:

> If a school principal claims that his or her school does not track and yet teachers in that same school can divide their classes based on the average abilities of their students, then clearly there is informal tracking taking place (Rees, Brewer, & Argys, 2000, p. 18).

The authors continued their argument stating that, in a study that aims to compare a tracked school to a non-tracked school, one should determine whether or not the non-tracked school is practicing informal ability grouping instead. In that case, the researcher is comparing formal versus informal ability grouping, rather than homogeneous versus heterogeneous classes (Rees, Brewer, & Argys, 2000).

Hoffer noted this phenomenon, stating that some schools that have ability grouping may not have a formal policy of grouping: "Students are informally encouraged to take or discouraged from taking certain classes by guidance counselors, teachers, and peers, even while admission to all classes is formally open to all" (Hoffer T. B., 1992, p. 212). The existence of informal ability grouping presents another problem in defining the practice. Since I could not

assume that schools without a formal ability grouping practice may have been grouped informally, it was assumed that if an author stated that a school did not have a formal ability grouping procedure, then the school did not practice ability grouping.

Debate over Ability Grouping

As stated earlier, the debate over ability grouping has been a constant feature of dialog among educational stakeholders over the past century and is still unresolved. Exploring three different meta-analyses on the subject made the why of this unresolved dilemma very apparent. Each study examined within-class grouping in secondary schools, and yet all three studies produced different results. Kulik and Kulik performed a meta-analysis on fifty-one studies and reached several conclusions. They found that thirty-six of the fifty-one studies reported more positive achievement effects in grouped classes (Kulik & Kulik, 1982). Using the index of effect size (ES), they calculated an average ES of 0.1, with a standard deviation of 0.32, meaning that in a typical class, performance of ability-grouped students was a tenth of a standard deviation higher than that of non-ability-grouped students. However, when they broke the studies down into individual characteristics, they reached other conclusions. The most surprising results came when the studies were characterized by the subjects in which students were grouped and the target group of the studies. Science grouping had the highest effect size of 0.18, followed closely by grouping of the combined subjects with an effect size of 0.15. When target groups were taken into account, Kulik and Kulik found that the Talented and Gifted group had an effect size of 0.33, as opposed to the Academically Deficient and Representative of Population groups, which both had effect sizes of 0.02. As a side note, they also found that before 1951, studies reported an average positive effect of 0.24, which

decreased from 1951-1960, 1961-1970, and 1971-1980 to 0.15, 0.07, and -0.01, respectively. Slavin's study reported slightly different results, and found that achievement effects were essentially zero for all grade levels (Slavin R. E., 1990). Finally, both of these studies were contrasted with an analysis by Lou, Abrami, Spence, Poulsen, Chambers and d'Apollonia. The authors found an effect size for student achievement of 0.17 favoring small-group learning, and stated that the overall results favored homogeneous grouping (Lou Y. , Abrami, Spence, Poulsen, Chambers, & d'Apollonia, 1996).

Kulik and Kulik (1982), Slavin (1990), and Lou *et al* (1996) each presented a meta-analysis with different results on the effect of ability grouping on students. Additionally, without knowing whether or not the schools who were the participants in the studies may have practiced informal ability grouping, it is difficult to know whether or not the results are trustworthy. Rees, Brewer and Argys analyzed Hoffer's data, compared the results to a study done by Betts and Schkolnik and found that in utilizing the same data source, Hoffer identified fifteen percent of seventh grade math students were not tracked by ability, while Betts and Schkolnik identified twenty-seven percent that did not track (Rees, Brewer, & Argys, 2000). Although not an enormous difference, it illustrates that even when the same data is analyzed, studies may come to different conclusions. It is easy to see why ability grouping is a practice over which its effects are still contested.

The studies discussed above pointed to an underlying problem with ability grouping practices. The practice of ability group tracking is enacted differently in most of the schools that use this practice. The research literature can again be called on to help achieve additional clarity on this point. Sørensen proposed that there are six main characteristics that differentiate

tracking programs: type of differentiation, selectivity, electivity, assignment criteria, inclusiveness and scope (Sørensen, 1970). The type of differentiation refers to whether students are grouped vertically or horizontally. Vertical grouping describes a system such as grouping students by grade who are the same age. Horizontal grouping describes those practices that employ curriculum differentiation, also known as within-track grouping. Selectivity describes the extent to which classes within tracks are homogenous; the more selective the system, the more homogeneous the classes. Selectivity can also express the gap between the groups, as the more homogeneous a class is, the wider the gap is between them. Electivity describes whether or not students are allowed to choose their own groups. Sørensen pointed out that the level of electivity depends on whether one is using vertical or horizontal grouping. Grouping vertically rarely relies on student input, whereas grouping horizontally does allow into consideration the student's choice occasionally. Assignment criteria describe how students are divided into groups or tracks. Electivity is one form of criteria, while I.Q. scores, past performance in classes and test results constitute others. Inclusiveness denotes the number of students assigned to the higher track. On the one hand, it would seem that the more students assigned to the higher track, the less restricted the program is. A high degree of inclusiveness does, however, exclude the small number of students assigned to the lower track and may add to the stigma of being in the lower group. Conversely, a school that that allows only a small percentage of students in the higher track increases the selectivity of the program. Scope implies the flexibility of the program-the higher the scope, the more time a student spends with the same group, thus reducing the heterogeneity of the classes.

Most tracking programs can be described using Sørensen's characteristics, and it is their high level of differentiation that makes it difficult to determine whether ability grouping benefits or hurts students. Gamoran studied whether or not these four characteristics affect the achievement gap (Gamoran, 1992b). In most instances, the differences were negligible. There were, however, several significant results. Gamoran found that the more mobility in the tracking program, the higher the math achievement. Additionally, the moderately inclusive systems had less between-track inequality and an increase in overall school achievement.

All of these variations contribute to the reasons why the ability grouping debate continues to plague the education community. In evaluating ability grouping, should we assess only one meta-analysis, considering how different their results can be? Are all tracking programs equivalent, or do their levels of selectivity, electivity, inclusion and scope determine which systems we should be examining? These questions, while important, were not the focus of this study. They should, however, be taken into consideration when researching. For the purposes of this study, any method of grouping by ability, whether it's called tracking or curriculum differentiation, wide in its scope or narrow, was considered in my research.

How Ability Grouping has been Affected

In this section, I wanted to set the stage for the research synthesis that is to come. Consider this an overview to create context. It starts with one simple recognition: ability grouping has changed greatly over time. It is possible that most of this change is a result of educational reforms over the past century. The Committee of Ten, a group appointed by the National Education Association in 1893, met to determine the future of secondary school curriculum. Only one question proposed to the committee concerned ability grouping, and

although history has shown that their recommendation was not carried out to a large extent, it was the starting point of this study. Documents that record the educational happenings of the early twentieth century U.S. showed a large push toward increased enactment of ability grouping in school classrooms. Did this correspond with the creation of the *Cardinal Principles of Secondary Education*, released by the Bureau of Education, or could it be attributed to the laws that created compulsory education? From 1919 until 1955, the progressive education movement was in full swing with the founding of the Progressive Education Association, a group whose purpose was to spread the movement. How deeply was ability grouping affected by this movement? Does part of the effect lie with administrative progressives and their ideals that revered a system of measurement and reliability?

These reforms just take us up to the mid-twentieth century, however, when a whole new world of education was introduced with the ruling of Brown versus the Board of Education of Topeka, Kansas (1954). A mere three years after the ruling, Sputnik could be seen streaking across the sky as the Soviet Union beat America in the Space Race. How did these two events, neither one strictly an education reform, determine the course of ability grouping? In what ways did desegregation affect the achievement gap? Were black students given the same chance as their white counterparts in the tracking system? And how pervasive was the outcome of the Space Race in America's system of ability grouping?

The last thirty years have shown several government-introduced education initiatives, two of which have greatly affect ability grouping. In 1983, the National Commission on Excellence in Education released a report entitled *A Nation at Risk*. The report brought attention to ability grouping and proposed several suggestions for solving the dilemma. Did

these suggestions narrow or widen the achievement gap? In 2001, the most recent government education initiative was released. *No Child Left Behind* (NCLB) has radically altered the course of education in America, especially in ability grouping practices. This initiative, a reauthorization of the Elementary and Secondary Education Act of 1965, seeks to create educational environments in public schools across the nation that will bring every student up to the same standards. To do this, persistent evaluation of schools making use of student assessments and school-wide progress reports make "adequate yearly progress" a watchword for all who come in contact with schools. What does NCLB reform mean for the future of ability grouping, and will it ultimately harm or help America's students?

All of these issues, questions, and notations point the reader toward the general timeline my research followed. Although all of these questions were not fully answered, I hope that I at least presented my findings on the subjects adequately and inferred what might be the effects. These questions were not meant as research questions, but as a guide, one that hopefully determined the path ability grouping has followed since the Committee of Ten to its future.

CHAPTER 3

METHODOLOGY

Historical Research in Education

Bybee defined historical research as "the process individuals use to identify significant ideas, events, persons and institutions of the past" (Bybee, 1982, p. 3). The key term, Bybee points out, is significant. There has been a great deal of change in education's past. The key is to define what information is needed, what is superfluous, and how to use the information to solve a problem. At the time of Bybee's publication, only 1.6% of the articles published by *Science Education* were historical in nature (Champagne & Klopfer, 1978 as cited in Bybee, 1982). A cursory examination does not suggest any change relative to these numbers, which speaks to the state of science education 30 years ago. Historical study of educational research is still seldom seen in graduate schools, as most students tend to spend their time on more "scientific" research. It can be difficult to see how history affects the present, but that is the job of the historical researcher.

Historical research in education can add to our understanding of present-day issues related to schooling and society. It allows current educators to critically examine education's past in order to more accurately assess current reforms. Additionally, historical research gives practitioners the chance to examine their own profession in order to modify behaviors, policies and systems. It is simple to claim that history cannot teach us anything about the present since every point in history has its own context. This does not have to be true, though. History does

not have to be a static set of facts; rather, it can be used to bridge the gap between confusion and understanding. The unique circumstances of the past may not always match with the present situation, but the actions taken can help guide current researchers to elusive answers.

Historical Methodology

Facts themselves have little meaning until they are recognized as significant and arranged in an order that tells a story (Bybee, 1982; Marius, 1989). It must always be remembered, however, that we can only tell a story about a documented past, artifacts that describe history. Additionally, we cannot assume anything about the documents we find. It is impossible to pose questions to the creators of historical data (for the most part) and so it is only what is on the surface that can be used as data. Finally, history is complex. Marius wrote that nothing from the past ever has a simple answer (Marius, 1989). Very rarely do events occur that are singular and without relation to other issues. To simplify the cause of a complex problem into the events of one point in history or the words of one person only exacerbates the issue.

Oversimplification of the past is not the only problem that historical researchers face. Every story has at least two sides to it. If, for instance, I were to look only at data gathered about ability grouping from *Gifted Child Quarterly*, the history of ability grouping would be seen to play out differently than only utilizing data from the *Journal of Educational Change.* By making sure that the sources I used were varied and representative, it was easier to claim that the conclusions I came to were more accurate.

Location of sources

In order to locate the sources required for this study, I used a broad definition of ability grouping, one that includes tracking. The descriptors I used in my search included: ability grouping, tracking, academic tracking, science, science education, ability and intelligence. Whether the ability grouping is done within-classes through curriculum differentiation or within-schools through different diploma tracks, the outcome is the same: students are separated by perceived ability. It is unrealistic to assume that I found every piece of information pertaining to my subject over the specified time period. Rather, it was my goal to find the most relevant data in order to produce a complete picture. It was also unrealistic to assume that I used every scholarly article published on my subject over the past one-hundred years. It was my hope that utilizing the methodology detailed below produced a sufficient and comprehensive history of ability grouping.

I performed comprehensive searches for sources pertaining to ability grouping, ability grouping in science, the achievement gap, and educational reforms over the past century. After searching, I first looked for relevant study titles. I then read abstracts to determine if the study would be helpful. Finally, I read and analyzed studies that pertained to my topic. If the data was relevant, studies referenced by the author were used as additional sources. Journals were also searched extensively, including *Science Education*, *Sociology of Education*, *The School Review, Review of Educational Research, American Education Research Journal, American Journal of Education, and Educational Evaluation and Policy Analysis Archives*. My sources, however, were not limited to these journals.

CHAPTER 4

CHARACTERISTICS OF TRACKING PROGRAMS

Throughout most of the 20th century, academic tracking was used in schools to differentiate students into what were believed to be appropriate learning categories. The effect of tracking on student achievement, however, is unclear, as apparent from a review of the current literature (Kulik & Kulik, 1982; Slavin R. E., 1990; Lou Y. , Abrami, Spence, Poulsen, Chambers, & d'Apollonia, 1996). Conflicting views about the effects of academic tracking are presented with each of these studies as well as others published in the past twenty years. Some studies find there is no effect at all. There are many factors that may contribute to this confusion. Sørensen postulated that, in addition to student individuality and social relationships, the varied characteristics of tracking programs are to blame (Sørensen, 1970). He identified six main characteristics that define tracking programs: vertical versus horizontal differentiation, electivity, assignment criterion, selectivity, inclusiveness and scope.

Before presenting an analysis of these characteristics, it is necessary to define some of the previously used language. Tracking, ability grouping and curriculum differentiation are all terms that tend to be used interchangeably. It is still not clear after a review of the literature which term should be used. It should be noted that tracking, curriculum differentiation and ability grouping will be used in this thesis to mean any type of plan that intends to *reduce* heterogeneity within classes, as concurrent with Slavin's definition (Slavin R. E., 1990).

Vertical and horizontal differentiation are two terms used to classify approaches to grouping students. Vertical grouping tends to reduce the amount of variation in characteristics responsible for the student's capacity for learning. Although students' capacity for learning can be attributed to many variables, vertically grouping students tends to increase homogeneity in characteristics such as physical health, mental maturity and emotional stability. A nearly universal example of this is age-grouping. At the age of five, children begin kindergarten and, with the exception of students held back from promotion, advance a grade level every year. Although age-grouping is an important type of differentiation that deserves further examination, it will not be included in this body of literature. Other types of vertical grouping such as physical health and mental maturity will be included in this analysis.

Horizontal grouping tends to reduce the amount of variation among students with regard to the knowledge and skills responsible for school achievement. Students who are horizontally grouped are taught in classes with a high amount of homogeneity in intelligence scores, standardized test scores, and past performance. This type of differentiation can also be a reflection of students' future plans. For instance, students who plan to attend post-secondary school may be assigned to the college-preparatory track, while students who plan to attend a technical school or work after high school would be assigned to the technical-preparatory track. Horizontal grouping exemplifies what most schools consider tracking or ability grouping, when students are placed in classes by perceived ability based on test scores. Additionally, most schools use several different forms of horizontal grouping, as well as at least one form of vertical grouping. Georgia public high schools use age-grade grouping (vertical), special classes

for disabled students (vertical/horizontal), gifted classes (horizontal), ability-grouped classes (horizontal), and different diploma tracks (horizontal).

The type of grouping used plays a part in determining the degree of electivity, or the degree to which a student's decision about which group they would prefer to be part of affects the group to which they are assigned (Sørensen, 1970). For instance, the specific form of vertical differentiation in which age is the criterion for group placement would never rely upon student choice. Horizontal grouping, on the other hand, very often allows for student input. Because both types of grouping are frequently used in the same setting, student choice may be a component of organizational differentiation, but not the only method of assignment. The perception of the degree of electivity within tracking programs may be inflated, however, by the illusion of student choice. Often times, students are guided along certain tracks according to peer relationships and teacher and administrator input (Gamoran, 1992b).

Electivity can determine the track in which a student is placed; there are, however, other assignment criteria that can be used to separate students into groups. This includes, but is not limited to, past performance in courses, current achievement level, or I.Q. tests (Sørensen, 1970). The type of assignment criteria that is used can influence the degree of selectivity within a system. Selectivity is measured as the amount of homogeneity produced by the class assignment: the more selective the system, the more homogeneous the groups. Selectivity can also be viewed as a measure of the gap between ability groups. In a highly selective system, the top ability group scores higher on tests and receives better grades than the same group in a less selective system. Gamoran noted this in his discussion of selectivity:

> By definition, highly selective tracking systems are elitist - they place high-achieving students together to form homogeneous classes. Tracking tends to be especially visible in highly selective systems, with high academic status awarded to the "cream of the crop." By emphasizing the top track at the expense of other tracks, selectivity probably magnifies between-track variation in students' educational attitudes and expectation … Moreover, highly selective tracking systems are often characterized by greater between-track variability in students' instructional experiences. Because teachers adjust instruction to student aptitudes, tracks that differ more in initial levels of student performance are likely to vary more in their instructional regimes and hence produce wider gaps in achievement (Gamoran, 1992b, p. 815).

Assignment criterion, electivity and selectivity are characteristics of systems used to determine group membership that affect one another. In a completely elective system, other assignment criteria would not be used and selectivity would be completely dependent on student choice. Conversely, in a system with a low degree of electivity, and therefore more dependent on other assignment criteria, the degree of selectivity would be high.

Sørensen's fifth characteristic relates to how tracking programs include or exclude individuals. Inclusiveness, as Sørensen defined, is "the number of opportunities assumed to be available at different educational levels" (Sørensen, 1970, p. 360). In programs with a higher degree of inclusiveness, more students will be assigned to or included in the higher ability track. Conversely, the lower the degree of inclusiveness, the fewer students there will be assigned to the higher ability track. Although more students assigned to the higher ability group may seem

to be a positive, varying degrees of inclusivity can be seen as a double-edged sword. Gamoran noted this in a study that analyzed Sørensen's characteristics as a determiner for student achievement:

> The larger the size of the college-bound track, the more salient it is likely to be - for those who are left out. For example, membership in a non-college program may incur greater stigma when it consists of the bottom 10 percent of the school's academic hierarchy than when it is the bottom 40 percent. Although an inclusive system is less elitist, it is highly visible and thus stigmatizes those left out of the preferred group. However, a system characterized by very low inclusiveness also probably raises the salience of the college track (Gamoran, 1992b, p. 816).

As related above, inclusiveness is a characteristic that can greatly affect students from all groups, but especially students in a low ability group. A highly inclusive system, although less elitist, can negatively impact low-ability group students' self-esteem and view of themselves. Inclusivity is a factor that needs to strike a delicate balance if it is to work to the advantage of both groups. Additionally, inclusiveness relies on the presence and degree of horizontal grouping. Since horizontal grouping can reflect, or even direct, students' future educational opportunities and career choices, inclusiveness is a key characteristics associated with this type of differentiation. Inclusiveness and selectivity have an inverse relationship. As inclusiveness decreases, selectivity increases. The fewer students included in the high-ability group, the larger the gap between the students in each group. As such, the converse is also true.

The characteristics previously discussed define what ability groups are present within high schools and which students are assigned to each group. Scope defines the extent to which students are assigned to the same tracks in different courses (Sørensen, 1970). A program with high scope is one that has a group of students spending most of their time in educational activities with the same group, whereas a program with low scope implies a high degree of change between groups. There have been several additions to this concept. In 1976, Rosenbaum coined the term "track mobility" to refer to the movement of students across tracks (Gamoran, 1992b). Oakes added to this concept by introducing the terms "extent," the proportion of classes tracks; "pervasiveness," the number of subjects tracked; and "flexibility," whether a single track is assigned for all subjects or for individual subjects (Oakes, 1985). The degree of scope can drastically affect a student's peer relationships. A less mobile system can produce more homogeneous friendships than one that allows for student mobility.

In order to assess tracking programs over the past century in public high schools, it is necessary to define a set of criteria by which each tracking system can be examined and evaluated. Using the above aspects of organizational differentiation as defined by Sørensen, the characteristics of tracking programs will be easier to measure and analyze. By utilizing this method, I will be able to more accurately evaluate the characteristics of tracking programs and trace the changes between past and present systems. In order to understand the complete history of academic tracking, I will begin by analyzing public education systems at the end of the nineteenth century, before the widespread implementation of the practice. A very different time for education, 1892 brought with it change in the form of the Committee of Ten, intelligence tests and the comprehensive high school.

CHAPTER 5

1892-1917: THE COMMITTEE OF TEN (A RECOMMENDATION WITH CONTRARY RESULTS)

In 1892, the National Education Association formed a committee to evaluate the state of secondary education. Charles W. Eliot, the then president of Harvard University, was chosen to chair the committee, made up of educators from around the nation. The work of the Committee of Ten as it came to be called was supported by nine subcommittees from different curricular subject areas. Two of the questions posed to the committee concerned the differentiation of subjects:

> Should the subject be treated differently for pupils who are going to college, for those who are going to a scientific school, and for those who, presumably, are going to neither? And at what age should this differentiation begin, if any be recommended? (NEA, 1894)

The Committee answered the first question unanimously in the negative, making the second question unnecessary. All ninety-eight members opposed the separation of college-bound and non-college-bound students into separate programs. It is obvious, however, that their recommendation did not guide the path of school reform as we see it today. To understand why secondary schools veered so far off from the Committee's ideas, the political, societal and economic influences of that time period must also be examined. At the outset, secondary

education was very different than what is seen today. Public schools were on the rise but still had low enrollment and the comprehensive high school, an extremely common institution at present, was still being developed. Additionally, there were several states that still had not adopted compulsory education laws.

There is very little evidence of widespread tracking during the first decade of the twentieth century. An examination of Sørensen's characteristics of tracking programs yielded very few results. Vertical differentiation was very common, typically in the form of age-grouping. Horizontal differentiation, however, was seldom used. More often than not, students were separated by school rather than by groups within the same setting. Neighborhood schools formed student populations based on where students lived. This differentiated students mainly by family income. In 1916, the superintendent of Cleveland schools, William Elson, wrote:

> It is obvious…that the educational needs of children in a district where the streets are well paved and clean, where the homes are spacious and surrounded by lawns and trees, where the language of the child's playfellows is pure, and where life in general is permeated with the spirit and ideals of America-it is obvious that the educational needs of such a child are radically different from those of the child who lives in a foreign and tenement section… (Cohen, 1968, p. 106).

Since horizontal differentiation was an uncommon practice at the time, both inclusiveness and electivity were characteristics that were very nearly absent, which meant that the curriculum

was the same for all students. Without different preparatory tracks, all students were, hypothetically, prepared for college. Additionally, because there were not separate tracks, students could not elect to take different courses.

The lack of horizontal differentiation also influenced selectivity, scope and assignment criterion. Without the differentiation, selectivity, or the measure of the gap between student groups, was null, as was scope. The only characteristic that could be truly evaluated was the assignment criterion. For example, the Vineland Training School for Feeble-Minded Girls and Boys enrolled students using family member recommendations. This could be considered a type of criterion. If, however, students separated by school were included as horizontal differentiation, then selectivity and scope could be considered high between schools, rather than within schools.

Three studies, released between 1897 and 1912, greatly contributed to the formation of ability grouping and described many elements of tracking systems that still pervade present-day public schools. The growth of the comprehensive high school contributed to the development of tracking. Simply put, the comprehensive high school was meant to serve all students. Prior to the comprehensive high school's origin, students were separated by schools. Students attended at the school closest to them, called neighborhood schools, or schools such as the Vineland Training School mentioned previously. Additionally, in areas where families of many different socioeconomic statuses lived, there were private institutions and city schools. Wealthier pupils attended the private institutions, while the poorer students attended the city schools. An article written by Paul Hanus in 1897 discussed what secondary schools should set out to accomplish. Hanus mentioned two very important ideas that would help to shape how

public schools approached curriculum differentiation and student education. The first explicitly stated on whom public high school should be focused:

> …Whatever functions may be attributed to the secondary school, none has been insisted upon by the community with greater emphasis and permanence than that the secondary school should disseminate the elements of general culture among the people. The public high school has been called "the people's college;" and this designation is by no means to be considered as an attempt [sic] to elevate the secondary school into a rank which it does not possess…it expresses, merely, the cherished expectation that the high school shall disseminate the beginnings of a liberal education-the elements of general culture--among those whose time and means will not permit a higher education (Hanus, 1897, p. 439)

Additionally, Hanus recommended that by grouping subjects around a pupil's vocational interest, he or she would be better prepared to achieve in said field.

The second tracking element originated in 1909, when Leonard Ayres authored a book entitled *Laggards in Our Schools: A Study of Retardation and Elimination in City School Systems.* One of the first extensive studies released on the subject, Ayres examined the "conditions, causes and remedies" of student retardation. The "retarded student" was defined as a student who was older than he or she should have been for the grade in which they were assigned. Ayres listed several causes for this, including: late entrance, irregular attendance, physical defects, race and sex. He also discussed several ways to "fix" the problem. Ayres touched on

the Batavia system, a method of dividing class time between individual and group instruction in order to tailor teaching to the individual ability of students. Introduced in 1898, this method had been seldom used in public schools. Ayres analysis of the plan led him to believe that the method sped up the slower pupils, rather than advanced the progress of the "quicker ones." His analysis of *the Cambridge plan*, however, allowed for the "rapid advance for the brighter pupils and special attention for the slower ones" (Ayres, 1909, p. 194). Using the Cambridge plan, teachers were assigned to help both groups of students. Brighter pupils could finish the curriculum in four years, while slower pupils could complete it in six. Whatever plan the schools adopted, Ayres' hope was that students would be able to learn within a flexible system that allowed for rapid advancement and appropriate remediation. Included in his recommendations were special classes for foreigners for the speedy acquisition of the English language and the differentiation of students with physical defects to better serve the school population. Ayres concluded his recommendations with a discussion on what the purpose of schools were:

> What is the function of our common schools? If it is to sort out the best pupils and prepare them for further education in higher schools, then the most rigorous system, with the severest course of study and the lowest percentage of promotions and the highest percentage of retardation is the best system. But if the function of the common school is, as the author believes, to furnish an elementary education to the maximum number of children, then *other living things being equal* that school is best which promotes and finally graduates the largest percentage of its pupils (Ayres, 1909, p. 199).

As Hanus' ideas of the comprehensive high school and grouping subjects around a pupil's interests grew in popularity, and as Ayres set the stage with a proposal for tailoring schools to promote the highest possible number of students, a French psychologist named Alfred Binet created a test that would become an integral part of America's tracking methodology in public schools. Without a method of sorting and classifying pupils, grouping was more subjective and typically based solely on teacher and administrator recommendations. In 1904, a new method was created by Binet called intelligence testing. By 1910, Henry H. Goddard, the then Director of Research at the Vineland Training School for Feeble-Minded Girls and Boys, proposed using Binet's test to screen new students for his school. He classified three groups of students: moron (I.Q. score of 51-70), imbecile (26-50), and idiot (0-25) (Goddard, The Kallikak Family, 1912). Goddard was a firm believer in eugenics, or the idea that humans should engage in planned selection through genetics, and concluded that the best way of dealing with persons of sub-normal intelligence was by segregating them from the general community or sterilization. His most famous book, *The Kallikak Family,* detailed a study of a family of below normal intelligence that he blamed on "bad stock" (Goddard, The Kallikak Family, 1912, p. 12).

Before intelligence tests were used at the Training School for Feeble-Minded Girls and Boys, home visits were used to confirm new students. In most cases, a plea was sent to the school to consider enrolling a pupil, at which time a home visit was made to ensure the pupil's suitability for the program. Whether the parents were married, how the parents lived and how the child was raised were all characteristics used to identify potential students. In the case of the Kallikak Family, Goddard traced Deborah's (a 19 year-old student with the intelligence of a

nine year-old) family history back four-hundred and eighty descendants. He found that there were, “thirty-six illegitimate [persons]… thirty-three sexually immoral persons, mostly prostitutes… twenty-four confirmed alcoholics… three epileptics… eighty-two died in infancy…three were criminal… [and] eight kept houses of ill fame” (Goddard, The Kallikak Family, 1912, pp. 18-19). While some of these characteristics were more legitimate than others, it is apparent that many of the “reasons” for a child’s sub-normal intelligence were a reflection of the beliefs of the time, and did not contribute to whether a child should be placed in the Training School. This was not, however, an uncommon method by any means. At the time, many education professionals believed that students were not created genetically equal: “Nature has not made children alike, nor is she fitting them for the same experiences and destiny” (Hartwell, 1907, p. 185). It is obvious today, however, that to determine a child’s future based on their “genetics” was a flawed method.

Binet’s intelligence test soon became a widespread method for sorting students into ability groups. An article written by H.C. Stevens, published just six years after Goddard began testing intelligence at the Training School, detailed a survey of children labeled “retarded.” Stevens classified retarded children into three groups: miscellaneous students were foreign-born children, children who entered school late or who were habitually absent, and “physical defectives”; incorrigibles were students who were “habitually lawless” and “harmful to the moral health of the school”; and the mentally subnormal (Stevens, 1916, p. 452). In addition to testing students’ I.Q.’s, nurses performed complete physicals on each student. Suggestions for providing assistance for these children included industrial training, placement at a state

institution for the “feeble-minded” or correction of the physical defects combined with placement back into the schools.

The Arlington Plan, a plan utilized in Arlington High School, Arlington, Massachusetts for grouping pupils by ability, was one of the first high school tracking plans to both be analyzed and use I.Q. testing. Each subject in Arlington had three classifications: honors, medium and slow. The plan was not, however, applied to chemistry, where the instruction was mostly individual. Each classification also had different grading methods. An A given in the honors group was denoted by A_1, whereas an A given in the slow group was denoted by A_3. The plan also implemented a college-preparatory track through the grades given to each class.

> The honor groups do more work in a given subject than the medium and slow groups. The latter two are expected to cover at least the minimum requirement for promotion. The work done by the medium and slow groups is about the same as that required of a regular class, based on traditional methods of selection. In order to earn promotion in any group a pupil must have an average better than D. Marks below B are seldom found in the honor groups and seldom above C in the slow groups. No college certificate is given in a subject in which the average for the year is below B_2, and a certificate is not guaranteed even with this average (Clerk, 1917, p. 27).

Because college certificates were not granted unless the average was B_2 or above, no student placed in the slow group for the entirety of their time in high school could attend post-secondary school. Instruction varied for each group. In the honors group, “the assignments

[were] longer, the work in the class [was] more rapid, less emphasis [was] given to instruction and drill, the instruction [was] less formal, the classes [were] larger, and the classroom work [was] more class work than individual," while the slow classes boasted opposite instructional tactics (Clerk, 1917, p. 27). The Arlington plan also differentiated courses according to student interest. These differentiated plans included: technical, college, general, commercial and household arts, which the author noted was a common practice in secondary schools by this time.

Policies for tracking by ability went from almost non-existant to extremely common in the twenty five years since the Committee of Ten's recommendation. The characteristics of tracking changed drastically in that time. Initially a system with almost solely vertical differentiation, no inclusiveness, electivity, selectivity, or scope within schools and assignment criterion based mostly on family recommendations, tracking became a common and structured system. First, horizontal grouping, which was, for the most part, previously absent, became widespread. Grouping subjects by ability and students' future vocational interests became common place by the second decade of the twentieth century. Inclusiveness, the number of students assigned to a college-preparatory track, is difficult to discern. It can be stated, however, that there was an increase in inclusiveness, given the characteristic was absent earlier.

The tracking systems proposed in the first twenty years of the twentieth century had limited electivity. Even before formal tracking policies, students were not given very much choice in which schools they attended, and then in which classes they were assigned. Selectivity, defined as the amount of homogeneity within classes, and scope, defined as the

extent to which students were assigned to the same ability group overall, were also difficult to determine. Since the Arlington plan made no mention of the degree of these characteristics, however, the degree of these attributes cannot be determined accurately. Assignment criterion was the characteristic that was most drastically altered. What began as students separated into different schools through family recommendations grew into school subjects separated into different classifications based on I.Q. scores. Binet's intelligence test became the main method by which students were sorted into ability groups. The characteristics of tracking programs changed greatly from 1893 to 1917. The period of time between 1918 and 1953, however, due to compulsory education laws, the Cardinal Principles of Education, and the Progressive Education Movement, would continue to alter the course of tracking and evolve the system into what can partially be seen today.

CHAPTER 6

1918-1953: COMPULSORY EDUCATION, THE CARDINAL PRINCIPLES, AND THE PROGRESSIVE EDUCATION MOVEMENT (OR HOW TRACKING TOOK OVER)

The United States Public Education system began to drastically change in 1918, when every state finally enforced compulsory education. Additionally, schools began to receive money from the state to begin transporting students, flooding schools with more students than which the United States had ever dealt. Approximately 21.3 million individuals of ages 5-20 recorded that they attended school in 1919 (U.S. Census Bureau, 1920). The 1920 U.S. census reported that approximately 85% of 14-year olds attended school, and although that amount dropped to a mere 20% by the time the students were eighteen, there were more children attending secondary school than ever before. Twenty-five years after the Committee of Ten met, a new committee was formed, in part due to the large increase in student-age population. Appointed by the National Education Association, members of this committee introduced the *Cardinal Principles of Secondary Education*, a document that addressed the need for reorganization in secondary schools. In addition, the objectives of secondary school and recommendations on the role of schools in differentiating students' learning experiences were included. The first section of the document established why secondary schools needed to be reorganized, citing the following: changes in society, changes in school population and changes in educational theory, which mentioned "individual differences in capacities and aptitudes among secondary-school pupils" (NEA, 1928, p. 2). The committee stated the seven objectives

of education as: “1. Health. 2. Command of fundamental processes. 3. Worthy home membership. 4. Vocation. 5. Citizenship. 6. Worthy use of leisure. 7. Ethical character” (p. 5). The third and largest section discussed how these objectives would be met, as well as the changes that would need to be made in order to meet the new demands of secondary school.

The “specializing and unifying functions of secondary education” very clearly stated the role of curriculum differentiation in the newly envisioned secondary school. The NEA Committee made several demands of secondary school. The first was that a wide range of subjects be offered to “to test and develop the many important capacities and interests found in pupils of secondary-school age” (NEA, 1928, p. 16). Capacity, a word mentioned several more times in this section, should be thought of, in this context, as perceived student ability. Exploration and guidance, the second demand, would be implemented so that the student could explore his or her own capacities. The third, adaptation of content and methods, demanded that teachers should adapt content and methodology to each student’s capacity. Flexibility of organization and administration secured “provision for maximum and minimum assignments for pupils of greater and less ability, and, under certain conditions, for the rapid or slow progress of such pupils” (NEA, 1928, p. 16). Lastly, the fifth point demanded differentiated curricula, the basis of modern tracking systems.

> The work of the senior high school should be organized into differentiated curriculums. The range of such curriculums should be as wide as the school can offer effectively. The basis of differentiation should be; in the broad sense of the term, vocational, thus justifying the names commonly given, such as agricultural, business, clerical, industrial,

> fine-arts, and household-arts curriculums. Provision should be made also for those having distinctively academic interests and needs (NEA, 1928, p. 16).

This description could just as easily have been written after examining secondary school classrooms today.

All of the previous demands detailed the specializing functions of secondary schools, all concerning academics. The unifying functions described the social functions of schools, such as the "social mingling of pupils" and the "participation of pupils in common activities…such as athletic games, social activities, and the government of the school" (NEA, 1928, p. 17). The committee also called for the continued establishment of the comprehensive high school as the standard secondary school in the United States. As discussed earlier, the comprehensive high school majorly contributed to the use of tracking within schools due to the massive variation of students served. By reinforcing comprehensive high schools in the public schools system, the committee helped to cement tracking as a means of differentiating curricula.

As the NEA Committee published the Cardinal Principles, a movement characterized by contrary views was born. In 1916, John Dewey published *Democracy and Education*, a book that helped to advance the ideas of the progressive education movement; in 1919, the Progressive Education Association (PEA) was founded with these ideas in mind to help reform American education. The PEA was based on several tenets that concerned the nature of education. Proponents of the movement believed that traditional education was meant to prepare students for future responsibilities, while progressive education emphasized the individuality of the student and linked experience and learning. A common theme from the PEA

was to learn by doing. The idea proposed by the progressives was that experiences contributed by traditional education were wrong for the student and that the quality of education depended on the quality of the experience. Ideally, the PEA hoped that by teaching students the skills required to solve problems, all subject material could be learned, rather than by teaching facts.

Additionally, two types of progressives were born from this movement (Labaree D. F., 2005). Pedagogical progressives subscribed to a romantic ideal of natural learning and emphasis on development. In other words, they wanted students to be allowed to explore, develop in their own ways and experience education at their own pace, a very individualistic approach. Administrative progressives, on the other hand, believed in a strictly practical education system. Working to reorganize the education system to better meet the needs of society and economy, administrative progressives were ultimately utilitarian. I.Q. testing was a very important part of this plan, and matching students' abilities with occupations that were suitable for their mental capacities was their ultimate goal. Their ideas were combined and published in the *Cardinal Principles of Secondary Education,* a document with strong ties to the administrative progressive movement (Labaree D. F., 2005).

At the base of all of these changes was a shift in the goals of secondary school. The Committee of Ten believed that all students should be educated and that schools should be devoted to the pursuit of knowledge. The Cardinal Principles created an entirely different atmosphere, a set of ideals radically altered from the academic minds of a quarter of a century past. Secondary school, rather than focusing on the education of all students at the same level, was to focus on the education of all students at differentiated levels. The primary goal of

secondary school became preparation for college, a level that only so many students could achieve, but one for which they were constantly striving, or falling.

The academic world exploded with papers and studies that detailed how schools grouped students by ability. Many used methods similar to *the Arlington Plan*. Students were separated into groups by some form of intelligence test or, in one case, by the *Chassell-Thorndike Graded Opposites Test* (one test), *the Pintner Survey Scale* (five tests), and *the Otis Group Intelligence Tests* (ten tests) (Glass, 1920). The test scores controlled how students were allowed to approach their schooling. There were even cases when students were recommended for full promotion by their previous school, only to be put into "study-coach" classes because of their test results. In order to validate the test results, Glass asked teachers to rank their classes and found that there was a one-hundred percent correlation between mental tests and teacher's judgment. Although an impressive correlation, one must wonder whether the teachers were told the test scores for each student before classifying them into groups, which may have skewed the results.

Other schools used a vast array of criteria to sort students into groups. H. H. Ryan described a program that used twelve pieces of information for each student:

> (1) general ability rank in class, as estimated by the elementary school, (2) chronological age, (3) mental age, (4) intelligence quotient, (5) reading rate score (Monroe), (6) reading comprehension score (Monroe), (7) arithmetic fundamentals score (Monroe), (8) height, (9) weight, (10) dentition age, as estimated by the school physician, (11)

> social age, as estimated by the elementary school, and (12) physical status, or condition of health, as estimated by the school physician (Ryan, 1923, p. 51).

These criteria were used to place students into group A (the most advanced group). Students with an IQ less than 120, pedagogical skills that placed below the median, social or physical retardation, or poor health as judged by the physician were excluded. Appleton High School, following the trend that many other schools had set, separated students into three groups: x (fast), y (medium) and z (slow) (Rasey, 1923). During the summer, the principal reviewed each student case, taking into account his or her tests and record. Rasey noted the low scope of the program in writing that a student may have been placed into the x group for math and the z group for language, but that this was not a common practice. He also mentioned the general mobility of the system, stating that sometimes mistakes were made but were corrected easily.

There was no electivity involved in these systems. Some schools operated under the guise of student choice, such as those described in a 1919 article written by Appell and Wolfson. They described a method for differentiating foreign language classes by placing all students into combined classes for three weeks, and then separating students based on teacher recommendations.

> Not only was the attempt made to determine whether a pupil could reasonably hope to succeed in the study of a foreign language, but each pupil properly qualified was guided and helped to make an intelligent choice. Prejudice and hastily made decisions had to

be combated. Those not qualified had to be convinced that it was to their advantage to postpone the study for the time being (Appell & Wolfson, 1919, p. 259).

It is very apparent that, although Appell and Wolfson said there was student choice involved, there was not. If a teacher-assigned below level student tried to enroll in a high level class, they were "re-directed" onto the correct path.

Differentiated programs, such as the different diploma types seen today, became much more common. Although not yet separated by a different diploma type, Ferris described a differentiated curriculum for students with more academic interests versus those with vocational interests.

A. A liberal curriculum designed for pupils whose major interest is in the biological sciences might have the following organization:

I. Core subjects: general science, biology, botany, etc.

2. Closely related group: chemistry, physics

3. Remotely related group: history, foreign language, or mathematics

4. Constants: American history and civics, English, physical education and hygiene

5. Electives: music, art, literature, history, etc.

B. A vocational curriculum:

I. Core subjects: vocational practices, etc. through projects

2. Closely related group: theoretical and technical subjects

3. Constants: same as in liberal curriculum

4. Electives: any subjects not required in Groups I, 2, and 3 (Ferriss, 1923, p. 266)

Another very apparent conclusion can be drawn by looking at the two tracks. One was very academically oriented, while the other contained very few purely academic subjects. Though Ferris did not mention the level of student electivity in his system, one must wonder whether it contained the same amount of student choice as Appell and Wolfson's system.

The selectivity of each system varied greatly. Some schools used I.Q. results to compare groups, such as one institute that calculated the average I.Q. results of the A group and the C group and found a difference of 33.9 points (Chamberlain, 1924). Syracuse city schools evaluated selectivity using the Minimum Essentials Test grade. The highest 12.5 percent was assigned to the A group, while the lowest 25% was assigned to the C group (Whitney, 1924). Atlanta schools had a much lower selectivity between each group. Students were separated into the obligatory fast, average and slow sections, then separated further into "F.F (the fastest of the fast), M.F. (the medium of the fast), or S.F. (the slowest of the fast)" (Lyman, 1925). These divisions were made for the average and slow groups, as well. Placement was determined from school records and I.Q. scores, and the author noted that, although there were ability differences in each group, they were much smaller than was normally seen. For example, the S.F. (slowest of the fastest) group had approximately the same ability as the F.A. (fastest of the average) group. Lyman also noted the relation between group placement and track choice. The academic track was made up of at least 50% fast students, approximately

40% of average students, and very few slow students. Alternatively, the vocational track contained 15% of the fast groups and more than 30% from the slow groups.

In 1927, E. E. Lewis did a review of the current tracking programs and classified these programs into seven categories. Special classes were for students who were deaf, crippled, subnormal or gifted, or had some other physical or mental defect. Ability grouping classes followed the classic x-y-z pattern. This type of grouping was most common and allowed pupils placed in higher ability groups to either accelerate their education or enrich their learning. Although acceleration had been used most often during the first twenty years of the century, learning enrichment had started to gain popularity. Remedial or supplemental instruction, also known as *the Batavia plan* (a system discussed earlier in this study), dealt with children who had a tendency to lag or skip schools, and was mostly treated with afterschool coaching. Clinics for problem children were used for students who had extreme difficulty with their schoolwork. In these cases, students were given remedial or supplementary instruction. Differentiated assignments were used in classes that did not homogeneously group students, but instead required some students, usually those deemed by the teacher to be learning at a more advanced level, to complete extra assignments. *The Winnekta Plan*, a system that had lessened in popularity over the previous ten years, did not group students at all and was not very common. Children were not ever held back, nor were they ever skipped ahead. Students completed both individual and group work throughout the day. All pupils were expected to master a set of knowledge and skills while, individually, each student would also contribute a project that exemplified his or her own talents. Additionally, the common method of grouping students by age fell to the wayside, as students were not promoted annually, but when they

had completed enough coursework to be promoted. Lastly *the Dalton Plan* was a system formed on three basic principles: individualized and self instruction, freedom and community living. Students did not move as a whole to each subject. Instead, a student moved to another class when he or she felt that the work was completed.

Most of the time, more than one of these methods was used in the same school setting. Special classes for students with disabilities, classes for gifted students, x-y-z grouping and special sessions for students with attendance issues were very common fixtures in America's public schools by the late 1920's. Although Lewis strongly supported *the Winnekta Plan* and *the Dalton Plan*, they were seldom seen after this time. Both plans exemplified individuality and allowed students to work at their own pace. Lack of structure, however, was the probable cause of the plans' disappearance. Administrative progressives, a massive force at the time, held rigidity and structure high on a pedestal. It is hard to imagine these grouping plans used in public high schools today, considering the amount of structure in place.

When the Great Depression occurred in 1929, school attendance dropped drastically. Without the money for books, transportation, shoes or even clothing, attending school became a luxury that many families could not afford for their children. Thousands of schools were closed, and high school student enrollment dropped from approximately 20 million to 4.4 million. Despite the major changes, tracking continued to dominate in high school. Educators continued to publish articles related to tracking, most of which continued the discussion of how best to group. A meta-analysis, written by Paul Rankin, detailed all of the various methods for tracking students. Although Rankin used the formal definition of ability grouping (homogeneous grouping of students with respect to one or more characteristics), he made sure

to note that the grouping of students with an I.Q. of above 130 (mentally gifted) or below 70 (mentally deficient) were termed special classes and not included in ability grouping. By 1931, tracking was almost a constant in public schools. Rankin wrote that, "Ability grouping [was] here considered basically as an extension of grading," a statement that exemplified just how much tracking programs had been incorporated into the public school system (Rankin, 1931, p. 200). In his evaluation of current assignment criteria, Rankin listed the three most common criteria as I.Q. test scores, teacher judgment, and other educational measures. Although most tracking programs were using I.Q. test scores as the primary assignment criterion, Rankin noted that many programs allowed for teachers to correct placement mistakes. For example, a study that recorded the number of reclassifications found that 77.8% of the students in group X were confirmed, while only 67.3% of the students in group Y and 63.4% of the students in group Z were confirmed. Interestingly, though I.Q. test scores were used as the preliminary grouping criteria, teachers were allowed to correct mistakes. This practice dwindled over the next decade as educational tests and I.Q. scores became the primary, and sometimes sole, criteria for separating students into groups.

Although most conceptions of ability grouping in schools at the time merely attempted to refine the practice of tracking, there were two important developments that impacted the future of curriculum differentiation: a recommendation for stronger differentiation between academic and vocational curricula and studies with data that began to reveal the disparity between black and white high school students. The first, a critical assessment of secondary school differentiation written by E. E. Cline, called for higher amounts of differentiation in curricular tracks and ability groups. He criticized the current differentiated curricula, arguing

that a year of separate instruction was not enough, and that the differences between academic and vocational tracks needed to be stronger (Cline, 1934). Furthermore, he argued that homogeneous grouping without a strict plan for differentiating assignments was pointless and created a hollow system without any real change. Cline recommended changing the coursework in each class depending on which track students pursued, not just including a few specialized classes. It was in this plea that Cline coined what would become very common terms.

> Out of these two problems came our first decision: to develop two distinct groups of courses--one for the academic (college-preparatory) pupils, one for the non-academic pupils. We began with the English course of study and divided it throughout into two paths: what we call the A path and the G (general) path. In the A path we left the classics, the grammar, and the composition usually considered profitable for those who are preparing for college. In the G path we eliminated the grammar, restricted composition to oral work and paragraph-writing, and in literature sought primarily to raise the level of appreciation in contemporary reading-news-papers, magazines, short stories, familiar essays (Cline, 1934, p. 433).

The college-preparatory track and the general track (today termed the vocational track) later became different diploma types, creating a new wave in tracking programs. Additionally, Cline pleaded for a difference in quality and not quantity between the two programs. The only

classes not tracked were art, music and physical education classes, extra-curricular activities and homerooms.

Cline's recommendation provided high schools with a way of changing not only the type of education students received, but also the degree with which they graduated. His idea would become increasingly important once schools were no longer segregated by race. Before that could happen, however, the differences between black and white high schools had to be exposed. Although only published in journals for black educators, studies citing data that revealed the disparity between black and white high schools began to get published. The results were telling, and showed that 55% of black students, as opposed to 36% of white students, were enrolled in the agriculture vocational track. By contrast, 30% of white students and only 16% of black students were enrolled in trade tracks (Wilkerson, 1939). Additionally, black high schools offering agriculture vocational tracks were nearly nine times that of black high schools offering trade vocational tracks. Although strong trends, it was difficult to compare any black schools to white schools, as most black high schools only enrolled students for one year and rarely were able to take advantage of state-funded vocational courses. None of this data was very surprising, however, when the amount of federal funding granted to each type of high school was considered. Each type of program had at least eight times more funding, and up to as much eighteen times the amount of funding, for white high schools. Purely academic tracks were rarely offered in black high schools. In Alabama, only 36% of black high schools listed preparation for college as an objective (Jackson, 1940). Additionally, 55.3% of schools offered a single curriculum. Emphasis in curriculum was slightly contrary to these numbers, however, with 62.2% reporting an academic emphasis and only 29.5% reporting a

vocational emphasis. Interestingly, only 26.2% of schools kept a record of intelligence or psychological scores; 77.1% of schools, however, used test results to determine the ability of pupils. Finally, 55.8% of schools used classification grouping. Tracking in black high schools, according to the data, had more of an emphasis on ability grouping between classes (i.e. different ability groups for each subject) than on offering different curricula for students. Even so, little more than half of the schools used grouping.

As New Deal economics began to rebuild America, student enrollment steadily rose. By the late 1930's, 6.5 million students were enrolled in high school, 2.1 million more students than there had been ten years earlier. Rarely mentioned before this time, the extreme ends of the ability spectrum, students with severe disabilities and gifted students, began to receive attention (Hockett, 1944; Woods, 1944; Newland, 1953). In 1942, Lenoir Burnside reviewed a plan utilized by a high school in New York City that provided special provisions for students who scored in the top one percent of intelligence tests. It should be noted, however, that these students were not necessarily the top of the class, but only the top one percent in test scores. The objectives for the "honors-work" students were:

> …to provide for our more gifted pupils a program of educational experiences which would help them to realize their full potentialities, teach them the habit of thorough study, stimulate intellectual curiosity, enrich and strengthen their personalities, and assist in preparing them to accept and meet with dynamic initiative the problems presented by a rapidly changing society (Burnside, 1942, pp. 275-276).

Although separated for academic subjects, the honors-work pupils were not segregated from their classmates for electives. The class used a modification of *the Dalton Plan* and allowed students to move at their own pace with little teacher intervention. Suggestions by the author included an honors program that spanned the entirety of the pupil's high school education and scholarships for gifted pupils who could not afford post-secondary school. One gifted student, unable to afford college, was granted a thirteen thousand dollar scholarship, a situation commented on by Burnside: "For a gifted child to be denied opportunity because of lack of money is a waste of our most valuable asset" (Burnside, 1942, p. 285).

Burnside's sentiment, a lone voice at the time of his publication, did not remain solitary for long. Alongside a burgeoning development of concern for the mentally gifted, tracking in science, previously only touched on as part of general ability grouping scenarios, began to grow. In 1933, 52.7% of secondary schools required only one year of science for graduation (Novak, 1943). Benjamin Novak, bringing attention to the low numbers of students required to take science, called for more funding, greater adaptation on the part of teachers to deal with students of lower intelligence and, most importantly, two versions of the science curriculum. One track made for the "layman" and one for the "specialist," they differed in many respects. The layman track would offer a more generalized view of science and a course that would combine physics and chemistry into one course, rather than require pupils to take one year each of the specialized sciences. Additionally, the layman track was meant to emphasize science from the consumer viewpoint (Havighurst, 1955). The specialist track, on the other hand, would focus on the specific concepts and work more toward preparing pupils for college sciences.

A new era was approaching. Stronger differentiation between academic and vocational tracks, focus on gifted and mentally disabled students, and tracking in science ushered public high school education into this new era. Murmurings of integrated high schools began to be heard across America, and as the United States prepared to enter a new phase in public high school education, the stage was set for a great deal of change. Horizontal grouping, though not as common in black high schools, had become extremely commonplace. The degree of inclusiveness was decreasing with the focus on gifted pupils, and the gap between ability groups, or the selectivity, continued to increase. Electivity, the measure of student choice, was almost an afterthought in many tracking programs. Even in systems that allowed students to decide, the choice was not always autonomous. Scope had become almost uniform, with most students being assigned to an ability group in the academic subjects, but not in electives. Although teacher opinion still held some weight in determining ability group placement, Binet's intelligence test and other educational measures continued to dominate, and even outrank, teacher placement. With desegregation on the horizon and the threat of communism growing, American education entered 1954.

CHAPTER 7

1954-1982: BROWN VS. THE BOARD OF TOPEKA, SPUTNIK, AND HOW GIFTED STUDENTS WON THE WAR

When thirteen Topeka parents filed a lawsuit against the Board of Education of the Topeka City schools in 1951 so that their children could attend the school closest to their homes, they could not have known the enormous impact their actions would have on the United States. The unanimous opinion from the Supreme Court Judges three years later ruled that separate was not equal, and stated that:

> Segregation of white and colored children in public schools has a detrimental effect upon the colored children. The impact is greater when it has the sanction of the law, for the policy of separating the races is usually interpreted as denoting the inferiority of the negro [sic] group. A sense of inferiority affects the motivation of a child to learn *(Brown v. Board of Education of Topeka,* 1954).

Although some schools initially resisted the order to integrate, the mound of resistance began to crumble. By 1957, through the intervention of President Eisenhower, even America's most resistant schools began to integrate. Although an enormous time in America's history, school integration was only one of two monumentally important events to affect the United States that year. As Sputnik, Russia's satellite, streaked across the night sky on October 6th, high

school education became a focus of parents, politicians and educators such as the country had never seen. These two events, seemingly unrelated, became essential to tracking procedures and public school policies.

It was an article published by *Life Magazine* in 1958 that jump started education policy makers into action (Crisis in Education, Part I: Schoolboys Point up a U.S. Weakness, 1958). With boldly stated wording on the front cover that read *Crisis in Education,* the article pitted Russian schools against American schools. Alexi Kutzkov, one of Russia's best and brightest, was shown playing chess after school, reading a complicated text in front of his classmates and working on material considered too advanced for America's schools. Stephen Lapekas on the other hand, a high school student from Chicago, joked with his fellow classmates about his ineptitude at math, swam eleven hours a week, rehearsed for the school musical for two months, and proudly boasted his one word a minute typing skills. Obviously, American schools were in a sorry state.

As a kneejerk reaction to the U.S.S.R.'s launching of Sputnik and the apparently obvious statement their victory in the space race made about U.S. education, the U.S. Congress passed the National Defense of Education Act (NDEA). An Act that provided increased funding to public schools, the NDEA ultimately failed to greatly affect public education. Concern over the spread of communism and the after effects of McCarthyism caused a mandate to be inserted into the document. In order to receive funding, the school had to complete an affidavit confirming their belief in the U.S. government. Although some schools complied, over 150 institutions protested, citing a violation of academic freedom. After four years of protestation, President Kennedy repealed the mandate.

The role models put forward in the images of Alexi Kutzkov and Stephen Lapekas changed U.S. public education in many ways. For one, the Progressive Education Association, a major force in the previous era, fell from favor. With both the changes in education and the political climate, progressivism no longer encompassed the current views on education. Kutzkov and Lapekas also pushed gifted students into becoming the main focus within public schools. How to educate them, funding for programs, and which children should be admitted were all questions that people strove to answer.

There were three primary ways to educate gifted children: enrichment, special grouping and acceleration. Additionally, there were two types of enrichment: simple and special group. Simple enrichment procedures attempted to make the work for gifted students more interesting or challenging without removing them from a varied ability classroom. An application of simple enrichment, for example, required all students to be assigned the same work, but with the gifted student going farther with the assignment or completing supplementary tasks. These tasks usually emphasized creativity, individuality, critical thinking, or extensive reading. Special group enrichment placed gifted students into a group for all or part of the school day, offering many of the same learning opportunities as simple enrichment. For instance, eight high schools in New York had, within the same setting, separate "schools for the gifted, where the abler pupils [were] segregated in practically all areas except health education, study hall, and lunch" (Havighurst, 1955, p. 326). Acceleration programs offered gifted students the opportunity to move at an appropriate pace for his or her ability, and allowed them to complete high school in a shorter amount of time. Acceleration also took two forms, with high schools utilizing either skipping plans or speed-up/special-progress plans.

Although both forms of skipping, they produced very different results. Skipping allowed students to move into a higher level grade, but then work at the pace of the new group. Special-progress plans allowed students to finish classes or grades at a more accelerated pace, and kept those students at that pace, regardless of their grade.

In Havighurst's analysis of the methods of education for gifted students, he found that the bigger cities utilized special grouping enrichment, while smaller cities and private schools used simple enrichment. Havighurst argued that because larger cities sometimes had tens of thousands of students, they were in a better position to use special group enrichment, while the smaller community schools had no need for such grouping. Additionally, he postulated that since families in large cities did not know each other as well, special classes for gifted students were not as "visible" to parents. In other words, the children who were not in special classes could be ignorant that the program even existed, whereas a smaller school community may have viewed the placement as showing favoritism. Havighurst also found that schools in upper-middle class suburbs were less likely to have special grouping programs, as they were "favored" communities with families of high socioeconomic status and 80-90% of their high school graduates going to college. They usually spent two times the amount of money per pupil in their schools than did the average community. The average I.Q. score was much higher than that of an average school community and the schools treated all students as if they were going to college. Parents in these communities wanted their children to have every possible educational advantage and had the means to accomplish it.

In addition to the focus on gifted students, science suddenly became intrinsic to America's success. In order to pull ahead of the Russians, United States students needed to be

well-versed in scientific concepts, especially gifted students. Dr. M. H. Trytten, a then member of the National Academy of Sciences, believed that success in science for pupils was strongly linked to ability grouping:

> But much more important is the atmosphere of the school, the common denominator of interest in the group with which the student associates, the social norms which exist in the group as it develops its own esprit de corps. Youth are social animals and tend to wish to be accepted as members of the social group in which they find themselves. If this group is reasonably homogeneous, and composed of young people most of whom have reasonably intellectual interests and goals which involve intellectual effort, the group as a whole can generate an atmosphere in which, by mutual stimulation, a high degree of intellectual effort may exist. If, however, the group is heterogeneous intellectually, the chain reaction fails. The group will still develop common denominators of interest but they will be of a wholly different and generally non-intellectual character (Kandel, 1959, p. 161).

Additionally, the BSCS released three different biology textbooks. Each of these textbooks was geared toward a specific ability group (Lisonbee & Fliegler, 1964; Metzner, 1964). Grouping pupils by ability, in general, a very common practice before, became an absolute necessity. This is apparent from Otto's study "Grouping Pupils for Maximum Achievement" when he said: "The title of this article is a dead give-away. These few words betray the educational climate in the United States at present" (Otto, 1959). Generalizations about what characteristics the superior

student versus the low-ability student possessed surfaced, typified in a list written by Myres in 1960. The superior student had a long attention span, liked mental independence, understood complex directions, was interested in many subjects, analyzed and discriminated, learned from their own mistakes, were energetic, self-directive, resourceful, able to relate school subjects to real-life situations, and capable of abstract thinking. The low-ability student possessed the opposite of all of the above characteristics (Myres, 1960). In addition to these developments, budgets for national testing programs were increased (Passow, 1960). The National Merit Scholarship was founded in 1955, only a year after the Brown decision, and only two years before Sputnik. This link between desegregation and Sputnik, however, was not coincidence.

An article written by Bruno Bettelheim in 1958 brought attention to the link between concern over desegregation and the grown need to educate gifted students.

> Since the Russians beat us, we, in turn, must beat somebody. If we could not beat the Russians, we could always beat our educators. So the cry went up: Our educational system has failed us; it's all wrong. A reform must be instituted immediately. We must have more and better scientists; our schools must produce them, and right away. Neither the press nor the man on the street who so loudly made these demands could say exactly what scientific discoveries had to be made or how education could possibly arrange for them. But such minor questions were not per-mitted to confuse the issue. Whatever doubts were raised were immediately drowned out in an even louder clamor for changes in our schools. In the deluge of demands, many of which were

> contradictory, the demand for special provisions for the gifted was especially strong and widespread (Bettelheim, 1958, p. 252).

When Havighurst found that suburban schools with a high socioeconomic status did not usually group students, there was a strong reason behind it. Bettelheim put it succinctly when he said: "There, the separation of nice white children from poor white and Negro children is accomplished by moving to the suburbs or by sending children to private or parochial schools" (Bettelheim, 1958, p. 253). His insistence that there was a connection between well-to-do families moving to the suburbs and the desegregation of schools seemed to be supported by Havighurst's findings. It was his reasoning behind the connection between desegregation, Sputnik, and education for gifted students that provided a logical bridge of events.

> School integration is compulsory not simply because integration is a social and a moral obligation. School integration is required by a mandate of the Supreme Court because, in states where the schools are segregated, educational facilities for Negro children are, in the judgment of the Court, neither truly equal nor adequate. This, then, is the charge officially leveled against segregated schools: they do not offer the children who attend them an adequate education. Public demands for better education for the gifted child are based on exactly the same charge: educational facilities are not adequate-this time, not for the Negro, but for the gifted child. Here we have the common denominator: lack of adequate educational facilities (Bettelheim, 1958, p. 255).

Desegregation meant, to many black students, a chance to finally learn at an equal level, in both monetary and experiential terms. For white public schools and, more importantly, white parents of students attending these schools, desegregation meant overall lower averages in academic achievement. With the public demands resulting from the political tension of the Cold War pressing down on schools for higher test scores and better educational opportunities, the answer came in the form of better educational opportunities for just the gifted. Thus, although there began to be equal opportunities for all students, there was also justification for better opportunities for gifted students. Ability grouping was the answer to all of this. If gifted students could be separated from their peers, either by grouping within schools or by grouping in private schools or public schools in the suburbs, then education could still be separate and equal.

Socioeconomic status, the biggest indicator for school achievement in the present day, began to garner attention. Pictures of students of different ability groups showed obvious differences. Low-ability students were sloppily dressed, with unkempt hair and a dirty appearance (Bettelheim, 1958). High-ability students, on the other hand, were well-dressed, with nicely combed hair and clean fingernails. These differences, although entirely superficial, became the clichéd version of these students.

> But again, why all this fuss about the gifted child? Have these so-called gifted children been winding up in coal mines? Have so few of them managed to enter Harvard, Yale, the University of Chicago, or the City College of the City of New York? Has the present system completely failed them? Do gifted children suddenly find that there is no longer

> a place for them in the universities? Are there suddenly fewer scholarships for them? Are our colleges un-willing to accept them before the age of eighteen, even though they are ready for college? (Bettelheim, 1958, p. 262).

The answers to all of Bettelheim's questions were answered with resounding no's, but to the climate of the United States public education system of the time, one would think that the questions had been answered in the affirmative. When considering the link between desegregation and ability grouping, the Supreme Court's decision should be remembered. They clearly stated that separate was not equal and that "segregation of white and colored children had a detrimental effect upon the colored children" (Brown v. Board of Education of Topeka, 1954). Although the court was only considering the race of the student, by replacing white with gifted, and "colored" with low-ability, it is easy to see that there was a very apparent link between desegregation and the sudden focus on gifted students.

In 1965, the *Elementary and Secondary Education Act* (ESEA) established Title I schools, or what are often referred to as schools for low-income students, as a school had to establish that at least 40% of its students came from low-income families as per a U.S. census. Additionally, the act established special classes for handicapped students. Both of these developments became very important for tracking. If a school was assigned as Title I, higher-income families moved, taking their children to other schools. Classes for handicapped students became national instead of local. This act, although well-known at the time, would become integral to education in its rewritten and reauthorized form of "No Child Left Behind."

Despite the link between desegregation and increased attention for gifted students, ability grouping did not change a great deal over the next decade. Increased funding for national achievement tests continued and tracking went on as it had before. In 1967, however, articles began to shine light on the sociological aspects of tracking. Studies that linked intelligence and ability to socioeconomic status became very common.

> The disparity between the performance of lower-working-class children, especially Negro children, and the performance of advantaged upper-working-class and middle-class children, Negro and white, is evident in the younger years, and increases in age and grade level so that from fourth grade on there is typically a difference of one grade or more in school achievement and as many as twenty points or more in intelligence quotient (Fowler, 1967, p. 77)

As more evidence for the link between socioeconomic status and ability level was piled on, Bettelheim's ideas from a decade earlier resurfaced. The idea of ability grouping as a form of segregation gained speed, and more studies analyzing ability grouping from a sociological point of view surfaced.

> Grouping by intellectual ability as it is generally practiced in our schools is often intellectual segregation. In each example of ability grouping cited [here], the opportunity of children of one intellectual ability level to associate with children of

> another intellectual ability level is significantly reduced, if not completely prevented (Johnston, 1967, p. 208)

Many of these authors purported that since students in different ability groups had little or no chance to interact with each other, they were essentially segregated. The main thought behind *the Brown v. BOE* decision was grounded in the idea that separate was not equal. With many, if not most, schools practicing some form of tracking or curriculum differentiation, many educational sociologists of the time believed that intellectual segregation was just as damaging to a child as racial segregation (Johnston, 1967). Johnston offered many ways to remove the widespread ability grouping from schools including: whole-class activities, multiple class grouping, grouping within a class, and individualization of instruction. Chief Justice Warren wrote that "Such an opportunity [to education], where the state has undertaken to provide it, is a right which must be made available to all on equal terms" *(Brown v. Board of Education of Topeka, 1954)*. It was the equality of opportunity that educational sociologists attacked in order to incite change in the public education system.

In 1970, Sørensen released his study on the characteristics of grouping programs, providing researchers a way to approach the ability grouping confusion. With his detailed analysis came research from other educators on the specific effects of ability grouping. Esposito published his 1973 research review of ability grouping studies and found two important themes. One, he found that homogeneous ability grouping essentially neither helped nor harmed students intellectually in general or specific groups (Esposito, 1973). In some studies that had significant numbers, there were slight gains that favored high-ability

group students, but these results were more than countered by the amount of studies that found negative impacts for average and lower-ability group students. Second, he found that the impact of homogenous grouping on development was negative. The inflation of self-esteem of high-ability students was counterbalanced by the stigmatization of average and below-average students.

The achievement gap, an expression used to describe the difference in achievement between the high- and low-ability group students, became a very common term (Ogbu, 1979); (Rosenbaum, 1980). Whether students were white or black, middle- or working- class, also became very common to note in studies. In 1976, Margaret Gordon released a study that detailed her analysis of the achievement gap between black and white students, and the gap between middle-class and working-class children. She found that achievement was often related to race and class, and that black students and working-class students were often the underachievers (Gordon, 1976). Additionally, Gordon found that black students underachieved more than their I.Q. scores predicted, while the white students overachieved more than their I.Q. scores predicted. In other words, even though both a black and a white student may have had the same I.Q., the black student would perform at a lower level than his or her score had predicted, while the white students would perform at a higher level than his or her score had predicted. This meant that the achievement gap between these two groups of students was even greater than many studies had estimated.

Despite what the Supreme Court had set out to accomplish with its ruling in the *Brown v. BOE* trial, desegregation failed to provide the equal chance for minorities that many had hoped it would. An article written by Cheng, Brizendine and Oakes in 1979 provided a

reflection of whether equality was a reality in public high schools. The authors purported that there were three reasons equality had failed in schools. The first was the false idea of a meritocracy. In public schools, it is assumed that all students have an equal chance at success and achievement. Due to the many differences students have in terms of home-life, however, not all students have an equal chance to succeed. The second reason equality had failed was because of the assumption that all education provides for upward mobility. In other words, that a public school education always gives students the opportunity and means to move up in socioeconomic status. The interaction of both of these concepts to provide high-status jobs only to a select group of students was the third reason the authors provided (Cheng, Brizendine, & Oakes, 1979). Tracking was linked to all of these, as it is a system based on meritocratic ideas that provides students with a path to their future professions. As schools continued to organize students according to race and socioeconomic status (accidentally or purposely), equality within schools continued to stay out of reach. Because of the stratification of races and classes that schools perpetuated, "separate but equal" persisted.

The thirty years following the desegregation of schools led to many changes in Sørensen's characteristics. Horizontal grouping in the form of grouping by ability was still very common across the nation, especially in schools in cities with more diversity. Electivity increased slightly everywhere with the invention of different diploma types, but the actual amount of student choice involved in tracking programs was not possible to discern. As discussed earlier, whether schools allowed students choice for academic group placement and how often student choice actually influenced their placement could not be determined, since the appearance of student choice did not always guarantee an autonomous decision (Gamoran,

1992b). The degree of inclusiveness in programs had greatly declined over the previous thirty years. Since gifted students were the focus of education more than ever before due to desegregation and the Space Race, fewer students had the option of pursuing the high-ability tracks or college-preparatory curricula. Scope, whether students were grouped in one, some, or all classes, varied depending on the school setting and the students involved. Summarizing Havighurst's findings from earlier, schools with a more diverse population kept students together as much as possible throughout the day (Havighurst, 1955). Schools in more suburban areas, however, were free to move students around during the day, as they had a more uniform group of students. Assignment criteria boasted the most marked changes. Teacher opinion was rarely used, if at all, in student group placement. Most often, high-stakes test results, intelligence tests and class performance determined student group placement.

It was science educators, however, that had suffered the brunt of America's clamor to provide gifted students with a higher quality of education. The reason the U.S. had "fallen behind" the Soviet Union in the first place was because of science education (Bracey, 2007). Although the blame would shift to language arts and math teachers in the coming years, science educators bore the brunt of the fault from 1954-1982. American education was at odds with itself. Academia screamed for equality, while schools kept trying to convince them that equality already existed. America was a nation at risk of losing one of its most convenient methods of sorting students, which is why, in 1983, the National Commission on Excellence in Education wrote a letter to the American people on the imperative for educational reform.

CHAPTER 8

1983-2000: A NATION AT (POSSIBLE) RISK

In 1981, T. H. Bell, the then Secretary of Education, created the National Commission on Excellence in Education in order to evaluate the quality of American education. The Commission was created as a result of Bell's concern about "the widespread public perception that something [was] seriously remiss in [America's] educational system" (1983, p. 5). Headed by David P. Gardner, the Commission delivered a report in 1983 that painted a desperate picture of the United States system of education. Just as desegregation, Sputnik and Alexi Kutzkov spurred the U.S. into action thirty years previously, so did the Commission's report. The document began with the deeply terrifying words "Our Nation is at risk; at risk of falling behind other nations and at risk of losing a war against ourselves that originated with Sputnik" (1983, p. 7). The Commission likened the state of education to an act of terrorism had it been committed by another country. Japan was building better cars, Korea was making steel more efficiently, and Germany was taking over the machine tool industry. America, the Commission claimed, was not just losing an industrial war, but a war for knowledge and skill. Learning was the "indispensable investment" that America would need to enter the information age (1983, p. 8).

In order to support their claims, the Commission listed over a dozen indicators they found during the research phase of the project. Some of these claims were:

- About 13 percent of all 17-year-olds in the United States can be considered functionally illiterate. Functional illiteracy among minority youth may run as high as 40 percent.
- Average achievement of high school students on most standardized tests is now lower than 26 years ago when Sputnik was launched.
- Many 17-year-olds do not possess the "higher order" intellectual skills we should expect of them. Nearly 40 percent cannot draw inferences from written material; only one-fifth can write a persuasive essay; and only one-third can solve a mathematics problem requiring several steps.
- The College Board's Scholastic Aptitude Tests (SAT) demonstrate a virtually unbroken decline from 1963 to 1980. Average verbal scores fell over 50 points and average mathematics scores dropped nearly 40 points.
- Average tested achievement of students graduating from college is also lower (1983, pp. 9-10).

They claimed that these deficiencies could not come at a worse time, as "computers and computer-controlled equipment [were] penetrating every aspect of our lives–homes, factories, and offices" and "One estimate indicate[d] that by the turn of the century millions of jobs [would] involve laser technology and robotics" (1983, p. 10). Unfortunately, not a single claim was backed by statistics. In fact, the document did not have a single chart, graph or reference to any survey or statistical analysis. The "deficiencies" that the Commission discussed were thin, though. The SAT, for example, underwent content changes in the mid-1970's, which may have been a cause for decreased scores. Functional illiteracy, as another example, was never

defined. What the Commission actually meant by the phrase was never clear. Additionally, whether or not any of the changes were statistically significant was not stated.

The second half of the Commission's report detailed the recommendations for education. Their recommendations were separated into five areas: content, standards and expectations, time, teaching and leadership and fiscal support. For high school content, the Commission recommended that all students take the same number of academic classes, but that college-bound students should also take two years of foreign language. Additionally, the Commission tried to set several concepts that should be covered in each of the subject areas. These concepts are very similar to the standards that most, if not all, subject areas are held to in education today. For standards and expectations, the Commission suggested that colleges and universities raise their academic requirements for admission and that standardized tests of achievement be administered at every major transition point in a student's education. The commission also suggested that the tests of achievement should be part of a nationwide system of state and local tests. Time recommendations included more homework, seven-hour school days, 220-day school years, and that placement and grouping of students be based on student achievement, rather than rigid adherence to age. Teaching recommendations included higher pay based on performance and 11-month contracts. Finally, the Commission suggested more federal government involvement for leadership and fiscal support.

The report concluded with a positive view on education, with the words "America Can Do It" (1983, p. 28). The Commission knew how to strike fear into the hearts of American citizens though, and despite not having any statistical evidence to back his claims, America followed through with their recommendations. Over the course of the next twenty years,

education evolved from a local system to a national one, conceivably with tracking at its heart. One development that started soon after the Commission's report was the idea that teachers had to earn gifted classes. In 1984, Finley described a suburban school that assigned its low-ability classes to new teachers. Suburban High School had teachers fill out request sheets for which classes they preferred to teach, but the process ended up as just a formality. The gifted classes were the property of veterans or those already teaching gifted classes. Unfortunately, this system created a very apparent disdain for teaching low-ability students, with some teachers going so far as to say that they were the "absolute pits" and teaching low-ability students was "defeating" (Finley, 1984, p. 235). Although it may have been more difficult or challenging to teach students in low-ability groups, believing that they were less important or not as rewarding to teach should not have been apparent. This phenomenon of expecting less from low-ability group students and receiving less in return was documented very thoroughly in *Pygmalion in the Classroom*, a book written by Rosenthal and Jacobson in 1968 (Rosenthal & Jacobson, 1968).

Over the next ten years, studies discussing the effects of ability grouping and how best to approach the age-old system grew in number. Gamoran, a huge proponent of the sociological aspects of tracking released several papers detailing the effects of ability grouping (Gamoran, 1989, 1992a, 1992b, 1993). Most of his papers discussed the social implications of tracking, and discussed whether or not tracking was an equitable system that should be continued. Additionally, many other authors published meta-analyses on the effects of tracking, including Slavin and Hoffer (Slavin, R. E., 1990; Hoffer, T. B., 1992). This explosion of papers was very connected to "A Nation at Risk," as many authors tried to either enforce or

refute the Commission's findings. Finally, in 1993, a report by Lawrence Stedman was published titled The Sandia Report and U.S. Achievement: An Assessment. Stedman concluded that, although the Commission's report was correct about some general trends, their analysis was flawed by errors, insufficient evidence and misuse of data (Stedman, 1994). SAT scores, for instance, had declined since 1966, but mostly due to changes in demographics and test design, concluding that the SAT decline was nothing to worry about. The Sandia report's final analysis was that scores had, when differences were accounted for, generally stayed the same. Unfortunately, the report received very little attention, and did not impact America nearly as much as *A Nation at Risk*.

Despite evidence to the contrary, tracking continued just as strongly as it had in previous decades. It was A *Nation at Risk*, though, that molded public education into much of what is seen today. Looking at public education in present-day Georgia, one can easily discern that many of the NCEE's recommendations were carried out. College-preparatory and technical-preparatory students both take the same amount of academic core classes, but college-preparatory students take at least two years of foreign language and more advanced science and math classes. Over the last seven years, the University of Georgia has raised its standards, so that each concurrent class is the most qualified that UGA has ever had. Standardized tests, once things seldom found in schools, have become nearly universal. In Georgia, students take the Criterion-Referenced Competency Test (CRCT) in elementary school and the Georgia High School Graduation Test (GHSGT) and subject-area End-of-Course Tests (EOCT's) in high school. A student in Georgia cannot complete their education without taking at least three difference types of standardized tests, not including the Scholastic Aptitude Test

(SAT), Assessment of Scholastic Skills through Educational Testing (ASSET) or the American College Testing Program (ACT). More homework and seven-hour days have been adopted, and although a 220-day school year has not been nationally adopted, several counties have tried it. Pay based on performance, a concept very recently introduced in Georgia by the governor, is to be voted on in the next few years, and tracking can be found in every level of public education, in several different forms. All of these changes have come about in the last twenty years since the publication of *A Nation at Risk*, leaving little doubt that its release was unrelated to these modifications.

Although the time between 1982 and 2000 was wrought with education reform, tracking changed very little. The only major change came in the form of standardized tests, with most, if not all, U.S. schools utilizing a similar assessment to determine student placement into ability groups. Despite more than thirty years of sociological arguments against it, tracking still had a great deal of support. This trend continued until the present-day, cemented into practice with another plea to the American people released in 2001. President George W. Bush's *No Child Left Behind* (NCLB) report placed into effect a system of education that held all students to the same standards, on the surface. Underneath, however, ability grouping became an integral part of education that schools could not change, even if change was desired. Additionally, the passing of the law tied federal funding to school achievement, something that had not ever been accomplished in America's past. NCLB brought America's schools into the current generation, with changes that have affected education in ways that have yet to be fully understood or evaluated.

CHAPTER 9

2001-PRESENT: NO CHILD LEFT BEHIND (THAT SHOULDN'T BE LEFT BEHIND)

When *A Nation at Risk* was released for the American people, its plea was simple. To provide the same exemplary education for all students, no matter their differences. It began with words meant to inspire action: "All, regardless of race or class or economic status, are entitled to a fair chance and to the tools for developing their individual powers of mind and spirit to the utmost" (1983, p. 7). Eighteen years afterwards, *No Child Left Behind*, expressed a similar sentiment: "…too many children in America are segregated by low expectations, illiteracy, and self-doubt. In a constantly changing world that is demanding increasingly complex skills from its workforce, children are literally being left behind" (NCLB, 2001). It has already been discussed that, while the National Commission for Excellence in Education may have had the best of intentions, *A Nation at Risk* failed to produce all of the positive effects that its authors envisioned. Many children continued to suffer from intellectual segregation in schools, and despite the recommended changes, school reform continued in the direction it had been heading: unequal schooling for students of differing measured abilities. *No Child Left Behind* promised to deliver what *A Nation at Risk* had not.

A reauthorization of the Elementary and Secondary Education Act of 1965, *No Child Left Behind* provided funding for Title I schools and handicapped students, as well as many other public education programs. A key difference between ESEA and NCLB, however, was the tying of federal funding to yearly school performance. The members of congress that reauthorized

ESEA believed that increased accountability was the key to improving schools, and in order to ensure improvement, funding needed to be linked to a yearly review. Each school would be measured according to state-wide yearly assessments. In the event that a school failed to educate its disadvantaged youth, assistance would be provided for the first year. If, after one year, the school still did not meet Adequate Yearly Progress (AYP), funding would be removed. Unfortunately, this form of punishment would cause more students to fail in meeting AYP. Finally, if after three years the school could not meet AYP, parents would be able to remove their children from the school and enroll them in a different institution using Title I funds. Additionally, the school would be restructured, which essentially meant that most of the teachers and administrators would be replaced. Schools that met AYP each year and continued to improve their performance would receive rewards in the form of additional funding. In addition, NCLB provided money for innovative programs, such as magnet schools and charter systems. Perhaps the most controversial stipulation of NCLB, however, was the goal of 100% proficiency for all students in math and reading on national assessments by 2014.

It has only been eight years since the passing of NCLB, but despite the short amount of time since its enacting, there have been many studies discussing the possible implications. Publications such as *Many Children Left Behind* (2004), a collection of articles that discuss NCLB as a new form of segregation and the effects of NCLB in the classroom, as well as articles such as Lance Fusarelli's discussion of the impact of NCLB cite many of the same criticisms (Meier, Kohn, Darling-Hammond, Sizer, & Wood, 2004). One very common critique is that a school can meet AYP on 28 out of 29 measures without meeting requirements. A school fails for the year if all measures are not met. Fusarelli details the plight of Durant Road Middle School, a school

with a reputation for excellence that failed to meet AYP because of failure to meet two out of twenty-nine targets, math and reading for ESL students. Additional criticisms involved the details of assessments. In 2003, 63% of Georgia's schools failed to meet AYP because fewer than 95% of the students in at least one racial subgroup did not take the assessment (Fusarelli, 2004). This requirement then, is more about student attendance than performance.

One very common critique is that NCLB reinforces a "teach to the test" mentality. Many teachers believe that the information they teach in the classroom is just drill and repeat, and does not teach the students concepts. Because of this, students may not be performing better in general, but merely on assessments. Additionally, NCLB is based on the idea that all states begin equally, as does each school. It is easy to see, however, that this is not true. One only needs to look at the difference between Georgia's Oconee County and Clarke County to understand. In addition, since schools are not funded purely through national means, district rules still apply. Local property taxes determine the amount of money each school system has to work with, and systems with higher-income families will benefit. Essentially, however, the main failing of NCLB is that it punishes schools that are unable to meet AYP instead of helping them. The concept is very similar to the debtor's prisons from the mid-nineteenth century, in which persons who could not pay their debt were put in prison until they could pay, which was impossible if they were in prison.

What does NCLB mean for tracking, though? Because of increased rigor and adherence to standards, it becomes more important than ever that schools track students. Heterogeneously grouping students can be a difficult, timely and costly process that may take years to fully develop (Watanabe, Nunes, Mebane, Scalise, & Claesgens, 2007). Because so

many teachers consider heterogeneous classes more difficult to teach, the transition from homogeneous to mixed-ability classrooms would be challenging. With the pressure to meet AYP, schools are not going to embark on an idea that will endanger their funding. Additionally, since students with mental handicaps or behavioral disorders are held to different standards than the rest of the school population, many students have been classified as needing "special" classes. Special education, another form of tracking, places students into classrooms without the option of receiving a normal high school diploma. These classes can be very helpful and, in some instances, essential for students. Unfortunately, many low-performing students are assigned to them in order to remove their scores from the school's AYP measure.

In 2007, a team of three researchers evaluated one school's effort to detrack chemistry classes. They found that, despite the four years it took to full integrate all chemistry students into one ability level of classes, the effort was successful. There were four beliefs that were instrumental to the success of the program, the most important of which was the "teachers' true belief in a developmental, as opposed to fixed conception of ability and intelligence" (Watanabe, Nunes, Mebane, Scalise, & Claesgens, 2007, p. 693). Oakes discussed this in *Keeping Track*, when she stated that tracking made sense for teachers who believed that intelligence scores and ability are fixed measures. In fact, the main concept behind tracking is that intelligence and ability may fluctuate, but never drastically change. Because these two measures do not change, it makes sense to arrange students into groups based their ability and intelligence quotient. Watanabe et. al. found that the chemistry teachers at Highlander High School did not believe that ability and intelligence were fixed, but rather were fluid and could

be developed. To teachers who resisted tracking, it was never "too late to learn" (Watanabe, Nunes, Mebane, Scalise, & Claesgens, 2007, p. 693). Low achievers, Nunes said:

> …don't believe that they can learn things easily so they need a teacher who believes in them even when they don't believe in themselves—someone who knows they can do it, and who doesn't let them go when they get stuck—who doesn't give up on them. It's that whole notion of teachers' ideas of intelligence influencing the classroom. The kids know I believe in them and I try hard to help them and I try hard to get them to meet me halfway (Watanabe, Nunes, Mebane, Scalise, & Claesgens, 2007, p. 694).

CHAPTER 10

CONCLUSIONS

Summary of Findings: Tracking the Changes

Over the course of the twentieth century, tracking went from being an almost nonexistent practice, to one of the foundations of the United States public education system. To understand the changes in tracking, we must look at the individual characteristics that Sørensen detailed: type of differentiation, electivity, inclusiveness, selectivity, scope and assignment criteria. In 1893, grouping students by ability was not common, whereas grouping students by age was a national practice. Over the next quarter of a century, ability grouping increased and grouping students by their vocational interests began. By 1954, both of these practices were widespread and only grew in frequency with the movement to desegregate schools. In addition to grouping students by ability and vocational interest, most schools provided special classes for gifted students, students with physical defects, and students with low intelligence quotients. With the publication of "A Nation at Risk" in 1983, ability grouping, as well as many other types of horizontal differentiation, was cemented in public high schools across the nation.

As tracking became a fixture in high schools, electivity was affected. At the turn of the century, students very rarely were allowed to choose which school they attended or which classes they took. Although the advent of different diploma types and curricula based on different vocational interests increased the amount of student choice involved in tracking

programs, the choice was never completely autonomous. In many instances, students were guided into a path, either by administrators, teachers, parents, or their peers. It is difficult to discern the degree of electivity involved in schools today, however. Many schools advertise student choice initially, but try and convince a student to change tracks later. For instance, public high schools in Georgia allow students to choose whether they would like to pursue a college-preparatory track or a technical-preparatory track. Although the initial choice may have been made autonomously, teachers and administrators may try and convince a student that a weakness in science, reading or math would be too difficult to overcome, something that would be easier if the student were pursuing a tech-prep diploma.

Selectivity and inclusiveness were also affected as horizontal differentiation increased in frequency. As mentioned previously, selectivity and inclusiveness have an inverse relationship. In 1918, as horizontal grouping increased, selectivity and inclusiveness were existent, but difficult to determine since many studies did not analyze the gap between groups. However, in 1954, with the increased focus on gifted students, inclusiveness decreased since fewer students were assigned to the top ability group. Selectivity, therefore, increased as the gap between the groups widened. Between 1970 and 1983, selectivity decreased as the gap between ability groups narrowed. In recent years, however, selectivity has risen again. Both selectivity and inclusiveness are dependent on the type of educational setting. In schools with a high degree of ability grouping, such as more diverse schools or schools in poorer areas, inclusiveness would be low and selectivity would be high. Alternately, in schools with low diversity and more money, inclusiveness would be high and selectivity would be low.

Electivity and assignment criteria are also tied together. In 1893, when students were not allowed to choose which school they attended, there was no electivity. Because there was no electivity, other assignment criteria had to be used to determine student placement, which at the time was family recommendations. When different diploma types became common, students were allowed to choose which diploma they pursued. Ability groups and special classes also existed at the time, however, and students were not allowed to choose which group in which they were placed. The assignment criteria at this time were most often standardized test scores, intelligence scores, or teacher recommendations. As the century progressed, teacher recommendations fell by the wayside and intelligence scores and standardized test scores became the most common methods of sorting students into groups. Today, standardized test scores most often determine group placement. Electivity, as discussed earlier, is a component of these programs, but is not necessarily truly autonomous.

It is my hope that the previous discussion of tracking characteristics over the past century has answered my two research questions: What are the characteristics of academic tracking practices in present day public high school science classrooms? and Do present day academic tracking practices have commonalities with practices that characterized classroom tracking earlier in the 20th century and, if so, what commonalities? The characteristics of present-day tracking programs have been stated. As to the second question, it is apparent that present day tracking practices share many commonalities with past practices. By analyzing tracking programs using Sørensen's characteristics, a clear path of change can be traced from 1893 to present day.

The only question that remains to be answered is my goal question: How can the history of ability grouping practices in public high school science classrooms inform present day practices in schools?

Implications

Over the course of my research, I began to realize that my goal question would be difficult to answer. The problem of how the past can inform the present is a very common theme in historical research. Why look at previous events if not to learn from them? The adage goes that history repeats itself; it therefore behooves us to learn from history. While this may be true, this is an oversimplification. Elements from the past reappear in many forms. It is how we interpret these similarities and differences that help us determine how the past is intrinsically linked to the present.

There are three main implications for this thesis. The first is the use of the analysis presented here to interpret current ability grouping practices. One of the major issues in determining the effects of ability grouping on students is the many different characteristics associated with grouping programs. By using a combination of Sorensøn's characteristics and the examples in this thesis of grouping practices throughout the last century, education researchers can more accurately describe and evaluate tracking programs. Tied to this is the second implication for this thesis: improvement, modernization and reevaluation of ideas based on past events. Tracking has been through many changes during the last century, with more yet to come. Tracking is not a perfect system, and perhaps not even a necessary one, but without consistent reevaluation of ideas and improvement and modernization of such programs, our mistakes will continue to plague us. My analysis shows that education is a

constant vehicle of reform. Although these reforms may not always work out for the improvement of student learning, innovation is endemic of public schools.

Tracking must be put to the test, for educators to either confirm or deny its value. My final implication concerns general understanding. By recognizing that ability grouping has many different facets, people can gain a more complete interpretation of tracking practices. It is not enough to evaluate tracking as a method of ability grouping. Researchers must explore all of the implications of tracking, including social and economic repercussions, to fully assess its effect on students, educators and schools.

REFERENCES

Appell, I., & Wolfson, A. M. (1919). Plan for organizing the entrants into the high school of commerce according to their attainments in scholarship, their capacities, and their aims. *The School Review, 27(4)*, 256-261.

Ayres, L. (1909). *Laggards in our schools: A study of retardation and elimination in city school systems.* New York: Russell Sage Foundation.

Bettelheim, B. (1958). Segregation: New style. *The School Review, 66(3)*, 251-272.

Bracey, G. W. (2007). The first time 'everything changed': The 17th bracey report on the condition of public education. *Phi Delta Kappan, 89(2)*, 119-136.

Brown v. board of education of topeka, 347 U.S. 483 (1954).

Burnside, L. H. (1942). An experimental program in education of the intellectually gifted adolescent. *The School Review, 50(4)*, 274-285.

Bush, G.W. (2001). *No child left behind*. Washington, DC: Office of the President of the United States.

Bybee, R. W. (1982). Historical research in science education. *Journal of Research in Science Teaching, 19(1)*, 1-13.

Chamberlain, E. (1924). Differences in ability which demand changes in subject-matter and method. *The English Journal, 13(9)*, 629-641.

Cheng, C. W., Brizendine, E., & Oakes, J. (1979). What is "an equal chance" for minority children? *The Journal of Negro Education, 48(3)*, 267-287.

Clerk, F. E. (1917). The arlington plan of grouping pupils according to ability in the arlington high school, Arlington, Massachusetts. *The School Review, 25(1)*, 26-47.

Cline, E. C. (1934). Differentiating secondary education. *The School Review, 42(6)*, 431-439.

Cohen, S. (1968). The industrial education movement, 1906-17. *American Quarterly, 20(1)*, 95-110.

Crisis in education, part I: Schoolboys point up a u.s. weakness. (1958, March 24). *Life*, pp. 27-35.

Esposito, D. (1973). Homogeneous and heterogeneous ability grouping: Principal findings and implications for evaluating and designing more effective educational environments. *Review of Educational Research, 43(2)*, 163-179.

Ferriss, E. N. (1923). Curriculum-building in the rural high school. *The School Review, 31(4)*, 253-266.

Finley, M. K. (1984). Teachers and tracking in a comprehensive high school. *Sociology of Education, 57(4)*, 233-243.

Fowler, W. (1967). Developmental science learning for disadvantaged children. *The Elementary School Journal, 68(2)*, 76-87.

Fusarelli, L. D. (2004). The potential impact of the no child left behind act on equity and diversity in american education. *Educational Policy, 18(1)*, 71-94.

Gamoran, A. (1989). Measuring curriculum differentiation. *American Journal of Education, 97(2)*, 129-143.

Gamoran, A. (1992a). Synthesis of research/Is ability grouping equitable? *Educational Leadership, 50(2)*, 11-17.

Gamoran, A. (1992b). The variable effects of high school tracking. *American Sociological Review, 57(6)*, 812-828.

Gamoran, A. (1993). Alternative uses of ability grouping in secondary schools: Can we bring high-quality instruction to low-ability classes? *American Journal of Education, 102(1)*, 1-22.

Glass, J. M. (1920). Classification of pupils in ability groups. *The School Review, 28(7)*, 495-508.

Goddard, H. H. (1912). *The Kallikak Family*. New York: The Macmillan Company.

Gordon, M. T. (1976). A different view of the IQ-achievement gap. *Sociology of Education, 49(1)*, 4-11.

Hanus, P. H. (1897). What should the modern secondary school aim to accomplish? II. *The School Review, 5(7)*, 433-444.

Hartwell, C. S. (1907). Promotion by subject and three-year courses. *The School Review, 15(3)*, 184-196.

Havighurst, R. J. (1955). Community factors in the education of gifted children. *The School Review, 63(6)*, 324-329.

Hockett, J. A. (1944). The mentally handicapped. *Review of Educational Research, 14(3)*, 217-223.

Hoffer, T. B. (1992). Middle school ability grouping and student achievement in science and mathematics. *Educational Evaluation and Policy Analysis, 14(3)*, 205-227.

Jackson, R. A. (1940). An evaluation of educational opportunities for the negro adolescent in alabama, I. *The Journal of Negro Education, 9(1)*, 59-72.

Johnston, A. M. (1967). Intellectual segregation. *The Elementary School Journal, 67(4)*, 207-212.

Kandel, I. L. (1959). Current issues in expanding secondary education. *International Review of Education, 5(2)*, 155-165.

Kulik, C. C., & Kulik, J. A. (1982). Effects of ability grouping on secondary school students: A meta-analysis of evaluation findings. *American Educational Research Journal, 19(3)*, 415-428.

Labaree, D. F. (2005). Progressivism, schools and schools of education: An american romance. *Paedagogica Historica, 41(1-2)*, 275-288.

Linchevski, L., & Kutscher, B. (1998). Tell me with whom you're learning, and i'll tell you how much you've learned: Mixed-ability versus same-ability grouping in mathematics. *Journal for Research in Mathematics Education, 29(5)*, 533-554.

Lisonbee, L., & Fliegler, L. A. (1964). The bscs and the slow learner. *The American Biology Teacher, 26(5)*, 334-337.

Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B., & d'Apollonia, S. (1996). Within-class grouping: A meta-analysis. *Review of Educational Research, 66(4)*, 423-458.

Lyman, R. L. (1925). The junior high schools of Atlanta, Georgia. *The School Review, 33(8)*, 578-593.

Marius, R. (1989). *Wrting about history*. New York: Pearson Scott Foresman.

Meier, D., Kohn, A., Darling-Hammond, L., Sizer, T. R., & Wood, G. (2004). *Many children left behind: How the no child left behind act is damaging our children and our schools*. Boston: Beacon Press.

Metzner, J. (1964). The gifted student program of the bscs. *The American Biology Teacher, 26(5)*, 341-344.

Myres, J. W. (1960). Identifying students of superior and low ability. *The English Journal, 49(7)*, 485-487.

National Commission on Excellence in Education. *A nation at risk: The imperative for education reform: A report to the nation and the secretary of education.* (1983). Washington, D.C.: United States Government Printing.

National Education Association. (1894). *Report of the committee of ten on secondary school studies*. New York: American Book Company.

National Education Association. (1928). *Cardinal principles of secondary education: A report on the commission on the reorganization of secondary education*. Washington: Government Printing Office.

Newland, T. E. (1953). The gifted. *Review of Educational Research, 23(5)*, 417-431.

Novak, B. J. (1943). Needed changes in science in the secondary school. *The School Review, 51(3)*, 164-168.

Oakes, J. (1985). *Keeping track: how schools structure inequality*. New Haven: Yale University.

Ogbu, J. U. (1979). Social stratification and the socialization of competence. *Anthropology & Education Quarterly, 10(1)*, 3-20.

Otto, H. J. (1959). Grouping pupils for maximum achievement. *The School Review, 67(4)*, 387-395.

Passow, A. H. (1960). Educating the gifted in the u.s.a. *International Review of Education, 6(2)*, 141-153.

Rankin, P. T. (1931). Pupil classification and grouping. *Review of Educational Research, 1(3)*, 200-230.

Rasey, L. C. (1923). A program arrangement for mental groups. *The School Review, 31(8)*, 608-611.

Rees, D. I., Brewer, D. J., & Argys, L. M. (2000). How should we measure the effect of ability grouping on student performance? *Economics of Education Review, 19*, 17-20.

Rosenbaum, J. E. (1980). Social implications of educational grouping. *Review of Research in Education, 8*, 361-401.

Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom*. New York: Holt, Rinehart & Winston.

Ryan, H. H. (1923). Grouping pupils for acceleration. *The Elementary School Journal, 24(1)*, 50-53.

Slavin, R. E. (1990). Achievement effects of ability grouping in secondary schools: A best-evidence synthesis. *Review of Educational Research, 60(3)*, 471-499.

Sørensen, A. B. (1970). Organizational differentiation of students and educational opportunity. *Sociology of Education, 43(4)*, 355-376.

Stedman, L. C. (1994). The sandia report and u.s. achievement: An assessment. *Journal of Educational Research, 87(3)*, 133-146.

Stevens, H. C. (1916). A survey of retarded children. *The School Review, 24(6)*, 450-461.

U.S. Census Bureau. (1920). School Attendance in 1920.

Watanabe, M., Nunes, N., Mebane, S., Scalise, K., & Claesgens, J. (2007). "Chemistry for all, instead of chemistry for just the elite": Lessons learned from detracked chemistry classrooms. *Science Education, 91*, 683-709.

Whitney, N. J. (1924). Ability grouping at syracuse. *The English Journal, 13(7)*, 482-489.

Wilkerson, D. A. (1939). The vocational education, guidance and placement of negroes in the united states. *The Journal of Negro Education, 8(3)*, 462-488.

Woods, E. L. (1944). The mentally gifted. *Review of Educational Research, 14(3)*, 224-230.