This is to certify that the

dissertation entitled

Changes In Math Attitudes Of
Mathematically Gifted Students
Taught In Regular Classroom Settings
From Fourth To Seventh Grade

presented by

Elizabeth Jane Hammer

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in School Psychology

Walter Hapkiewicz
Major professor

Date Jan 12, 2002

CHANGES IN MATH ATTITUDES OF MATHEMATICALLY GIFTED STUDENTS TAUGHT IN REGULAR CLASSROOM SETTINGS FROM FOURTH TO SEVENTH GRADE

By

Elizabeth Jane Hammer

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Counseling, Educational Psychology
and Special Education

2002

ABSTRACT

CHANGES IN MATH ATTITUDES OF MATHEMATICALLY GIFTED STUDENTS TAUGHT IN REGULAR CLASSROOM SETTINGS FROM FOURTH TO SEVENTH GRADE

By

Elizabeth Jane Hammer

Despite concerns expressed in the literature for the abatement of enthusiasm for mathematics if gifted students are not challenged, there is a dearth of empirical studies about mathematics for gifted children below age twelve. The current study was designed to answer the question of whether gifted math students' attitudes declined when they were not given access to accelerative options during the later elementary years and into middle school. It was hypothesized that a decline would occur as a result of a lack of challenge. Because the body of literature on attitudes toward math is generally couched as a gender issue, gender differences were also examined.

Ample empirical evidence suggests that most children do experience a decline in enthusiasm for mathematics, which accelerates precipitously at entrance to middle school. The current study found a decline for the gifted students, but it was less severe than that found in the general population, and was off-set by increases in other mediational factors --factors which also declined for the general population during these years. The hypothesis--that gifted math students may experience a serious decline in their attitudes toward math when there is no acceleration of their math curriculum during the later elementary and early middle school years--was not supported.

However, some patterns emerged which suggested that concerns in the literature are not completely misplaced. Teacher's attitudes toward the gifted student were less

favorable than would be predicted by the students' abilities, and paralleled students' attitudes toward the psycho-social value of success in mathematics and of the usefulness of mathematics in their lives. The gifted students failed to gain confidence in their ability to do math, and their anxiety generally increased as they moved through the grades. Both patterns raise questions about the wisdom of failing to provide challenge for these students as soon as their abilities can be identified, both for the fulfillment of the individual child, and for the psycho-social climate which might develop if success and talent in math were more publicly acknowledged. Anecdotal evidence in the literature for elementary aged students suggests they respond with enthusiasm and increased confidence to accelerative options. Finding the opposite pattern, albeit still at a more favorable level than for the general population, suggests harm may be resulting from lack of acceleration.

Gender discrepancies were similar to those in previous studies. Boys held much more stereotypical views of math than girls, for all age and ability levels. For other attitude variables, at most grade and ability levels girls held a less favorable view (not necessarily reaching significance) than boys. Girls increased their attitudes until fifth grade, when a decline began for the general population and middle-high ability high girls. However, high-high ability girls developed attitudes which were more favorable than the high-high ability boys' when they reached seventh grade. This finding, previously unreported, suggests some possible interventions which might be helpful to less able girls, to maintain or increase their fifth-grade favorable attitudes.

An additional outcome was establishing an adaption of the Fennema Sherman Math Attitude Scale (1976) as a valid measure for use with high ability later elementary and middle school students.

Dedicated to
Carolyn Mahalak, for her help at exactly the right times;
Barbara Cherem, for her ongoing encouragement and modeling;
and especially to Frank Hammer, my husband, whose
unfailing support, patience and belief in me through the years
has made this all possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Introduction

Overview

The gifted education literature has favored the acceleration of precocious mathematics students, beginning as early as it can be determined that students have unusual ability in mathematical thinking (Waxman, Robinson and Mukhopadhyay, 1996; Van Tassel-Baska, 1981, 1992). Since 1971, the vast majority of systematic acceleration in mathematics in the United States has occurred at the secondary level, stimulated in large part by the talent search model instigated by Julian Stanley at Johns Hopkins University, and replicated at other centers throughout the U.S. (Stanley, 1996). The Study of Mathematically Precocious Youth (SMPY), a longitudinal study of students who scored in the top 1% of the talent search cohort (which is, by definition, the top 5% of the general population) has created a wealth of information about the characteristics as well as the most effective options for development of mathematical talent (Benbow and Arjmand, 1990; Benbow, Lubinski and Suchy, 1996). The acronym SMPY has come to characterize the talent-search model at all age levels.

One of the major findings of these studies is a persistent and pervasive discrepancy between males and females in math ability, with statistically significant and increasingly greater differences favoring males as ability levels increase (Fox, 1976; Lubinski, Benbow and Sanders, 1993). Coming as it did at the peak of the cultural sensitivity to feminist issues, this otherwise arcane finding helped stimulate a surfeit of studies of gender discrepancies in math ability, as well as attitudes about mathematics (Eccles, 1985; Fennema and Peterson, 1985). These studies found the academic differences at most ability levels favor boys to a modest and statistically insignificant degree. However, the issue of access to intense training of mathematics, and to the career trajectories which would logically follow, continue to be viewed through the lens of gender equity (Eccles, Wigfield, Harold and Blumenfield, 1993; Hyde, Fennema, and Lamon, 1990; Hyde, Fennema, Ryan Frost and Hopp, 1990). Thus, most of the empirical evidence about the relationship between attitudes toward mathematics and mathematics achievement is reported as an issue of comparisons between genders for secondary and post-secondary students.

The knowledge base on precocious mathematics students below the age of 12, however, is far more scant, and even fewer studies deal with affective or attitude issues among this population. Recently, there have been some initial reports of programs modeled after the SMPY model being used for elementary and even primary aged students (Lupkowski-Shoplik and Assouline, 1993; Miller, Mills and Tangherlini, 1995; Waxman et al., 1996). These reports have focused on academic outcomes, which have been positive. As was true for older students, acceleration is an option which seems to work extremely well, and to be positively embraced by the younger students.

The issue of acceleration has, inevitably, become entangled with the politics of school equity and reform. While it is possible that students could be accelerated without separate grouping, in the real world of school, acceleration is nearly always implemented through some sort of ability grouping, and this issue has created the greatest resistance to accelerative options at all levels, K through 12 (Kulik and Kulik, 1991, 1997). Thus, ability grouping has frequently been abandoned in order to preserve equity (in appearance if not in fact, [Loveless, 1999 b]), to enhance all students' access to high caliber instruction, and to avoid the presumed negatives of labeling of students (Oakes, 1985; Slavin, 1987). The individual differences of students are presumed to be adequately addressed through differentiating the curriculum with techniques like enrichment, cooperative learning, interdisciplinary thematic instruction, etc. (cf., George, Stevenson, Thomason, and Beane, 1992; Tomlinson, 1999).

It is clearly documented that gifted children strongly dislike having to wait while other students catch up, and it is also common knowledge that they learn to cope by hiding their abilities rather than risk being seen as "brains," "geeks," etc. (cf. American Association for Gifted Children, 1978; Delisle, 1987; Galbraith and Delisle, 1996; Higham and Buescher, 1987). This is an issue of particular concern for girls, in light of the negative discrepancies shown for females in the gender comparison studies of math achievement and attitudes. Kerr (1994) describes how bright girls, ironically using their giftedness to make a healthy adjustment to their psycho-social environment, actually create internal barriers to achievement. Thus, if achievement is not valued and if challenge is absent, the accommodating psycho-socially alert girl will quickly adjust by displaying modest achievement.

Indeed, the gender discrepancy studies show that girls begin a relatively serious downward spiral in attitude toward mathematics, beginning in the sixth grade (Meyer and Koehler, 1990; Wigfield and Eccles, 1994). Gifted boys, who may have spent critical elementary school years in classes where their more socially conscious girl classmates hide or discount their math abilities, may thus confirm the culturally embedded belief that math achievement is a male domain. Many bright students of both genders may eventually develop the value that doing your best is less important than fitting in--and the resultant attitude toward math may decline enough that it impacts on eventual math motivation and achievement (Ablard and Tissot, 1998). Particularly "in the case of mathematical talent, advanced ability can be actively discouraged if children are forced to repeat, over and over, low-level skills they have mastered long before" (Waxman et al., 1996, pg. 5). Challenge can, in theory, be provided through curricular enrichment as well as acceleration. However, it is a rare mid-elementary classroom where within class enrichment (thus avoiding grouping concerns) constitutes a significant, or even noticeable, part of a student's school day (Delcourt, Loyd, Cornell, and Goldberg, 1994; Westberg, Archambault, Dobyns, and Slavin, 1993). In addition, acceleration is an option which has been shown to be effective in study after study, though as already noted, it is relatively rarely done.

Statement of the Problem

Though it is usually thought of as a primary value of the medical, not educational, profession, the mantra of "First, do no harm" is a salient one which has not been empirically examined. Are educators, though well intended, naively creating the failure of many students to achieve at high levels mathematically by failing early on to acknowledge and challenge the students who have the strongest talents in this area and would bring the strongest motivation to it? Fennema and Sherman (1976) defined several clusters of attitudes believed important to success in higher mathematics. Those were enthusiasm and persistence at complex problem solving, perception of self as a successful mathematician, belief in the usefulness of mathematics, confidence in one's competence at challenging mathematics, desire to acquire advanced levels of mathematical skill. Are these attitudes present in young precocious mathematicians to a greater degree than in their age-level peers? Is it possible that these attitudes are not well supported in a milieu which offers low academic challenge to the young mathematician, by preventing access to more advanced levels of skill development? If not supported at a young age, might these attitudes atrophy by the time the curriculum shifts from arithmetic to algebra? The critical question is: Could failing to provide adequate challenge to elementary aged precocious mathematics students contribute to a decreased level of enthusiasm and interest in mathematics? This study empirically examined this issue.

Research on this issue has important implications for educators making decisions about the pacing of mathematics instruction for gifted students. If math attitudes of the gifted decline in the absence of appropriately advanced curriculum, there would be solid empirical evidence that "regular," or even enriched, instruction of precocious math students may actually be harmful to the very attitudes and predispositions expected to motivate these students to study advanced mathematics. Conversely, if little change of attitude is found, or if gifted math students continue to sustain stronger attitudes than their age peers, then the concern for the eventual deterioration of involvement in advanced mathematics may be ameliorated. Though not a gender issue *per se*, the issue also has important implications in addressing gender differences in this regard. Previous findings that show discrepancies in math attitudes with girls showing more severe declines during the middle school years. The question of gender discrepancies in math attitude is unexamined for young high ability students. Empirical documentation of gender differences in attitude changes for precocious mathematicians during the elementary years will also add to the small number of studies examining this issue.

Review of Related Literature

High Mathematics Ability

The SMPY Studies

In the early 1970's, the Study of Mathematically Precocious Youth (SMPY) initiated an efficient and effective way of both identifying and programming for students who demonstrate precocious mathematical reasoning ability (Stanley, Keating and Fox, 1974; Fox, 1981). Using a Diagnostic Testing-Prescriptive Instruction (DT-PI) approach (Stanley, 1978), the SMPY model invited students of junior high age who scored above the 95%ile on a grade level standardized test to participate in the College Board's Scholastic Aptitude Test (SAT) (or, more recently, the American College Testing Program- ACT) under the same conditions and scoring norms as the more typical cohort of college-bound high school juniors and seniors. Students who did exceptionally well (i.e., better than the mean for the high school cohort) on the Mathematics or Verbal portions of the out-of-level exam were deemed to be candidates for radical acceleration, fast-paced or high intensity courses in their area of strength, mentoring, and guidance for specialized educational planning (Benbow, 1986; Stanley, 1991; Stanley, 1996). The SMPY model has been implemented nationally through regional talent searches for nearly three decades, with considerable impact on the development of options at the secondary level for students of high ability (Assouline and Lupkowski-Shoplik, 1997; Boatman, Davis, and Benbow, 1995; Olszewski-Kubilius, 1998; VanTassel-Baska, 1996).

Longitudinal studies have followed several cohorts of the most highly talented students--typically, students who scored above 700 (+2 s.d.) on the SAT-M before age 13 (Benbow and Lubinski, 1997; Benbow, Lubinski and Suchy, 1996; Brody and Blackburn, 1996; Lubinski and Benbow, 1994; Lubinski, Webb, Morelock, and Benbow, 2001; Stanley, 1996). The most highly organized of these, planned for a total of fifty years from its inception in the 1970's, was originally based at Iowa State University and is following students involved in SMPY's early years. (The study center has recently moved to Vanderbilt University.) These students, who now number several thousand, are providing a rich and increasingly insightful data base on both academic and psycho-social issues for the adolescent and adult development of individuals having extreme academic ability.

Several significant findings are revealed through the research on these students, recently summarized by Benbow and Lubinski (1997). First, it is possible to identify with a high degree of accuracy at age 13 the students who have the greatest potential to become the nation's great mathematical and scientific achievers, and students who eventually enter careers in the math/science pipeline are disproportionately found among those who score at the highest levels (top 1%) on the SAT-M at age 13. The abilities thus identified are robust, but seem, nonetheless, to flourish more fully through a planned program of specialized experiences. Most students reported the simple fact of being identified as gifted (through the talent search) as a positive experience because it validated their own self perceptions, sometimes for the first time. However, a few reported that it created more difficulties with grade level peers and, occasionally, local teachers. With regard to accelerated experiences, Lubinski et al. (2001) report "this special population strongly preferred educational opportunities tailored to their precocious rate of learning...with 95% using some form of acceleration to individualize their education" (p. 718). The participants occasionally expressed regret at not having more opportunity to accelerate, and noted that they believed the acceleration was helpful in their social and emotional development. Irrespective of the mode of acceleration, the programs produced generally favorable results and high student satisfaction with their experiences, both in the short and long run (Benbow and Lubinski).

The Elementary Years

Despite this rich data base about gifted children above the age of 12, empirical studies about mathematically talented children before the age of 12 are very limited, and there is hardly any information on attitude or affective variables in this group. This is no doubt a reflection of the impact of the SMPY model on selecting for math precocity at junior high, and although retrospective information has been gathered for some of the older groups, there is a far more limited understanding of the effects of psycho-social variables before junior high on mathematics talent development.

In 1991, based on the success of the junior high age Talent Search model, institutionalized implementation of SMPY was expanded to the elementary level (Assouline and Lupkowski-Shoplik, 1997; Lupkowski-Shoplik and Assouline, 1993). The initial results are promising, and have lead to the implementation of identification programs like the Midwest Talent Search for Youth (MTSY) based at the Center for Talent Development at Northwestern University, and the Belin Elementary Student Talent Search (BESTS) based at the University of Iowa. Both MTSY and BESTS use tests designed for and normed on eighth and ninth grade populations with third, fourth and fifth graders. To date, there is very little information on either psycho-social variables affecting this group, or on the implementation or effectiveness of programs evolving from these, and related, efforts.

Two exceptions are studies of the Model Mathematics Program (MMP) and the Center for Talented Youth (CTY) program (Daniel and Cox,1988; Miller, Mills and Tangherlini, 1995). Modeled on the SMPY DT-PI approach, these programs were piloted in the mid-eighties and sought to develop a blend of enrichment and accelerated options for students from second through sixth grade (later extended through high school). Using the School and College Ability Quantitative Test (SCAT-Q)[1] (instead of the SAT-

M) as the out-of-level test, fast paced instruction was planned through an Individualized Educational Planning (IEP) process for those who scored above the 65%ile on the out-of-level test norms. Those scoring below the 65%ile were involved in significant enrichment to supplement the normal mathematics curriculum. Reported results on these relatively small groups, as in the longitudinal studies of older cohorts, were nearly all positive (Daniel and Cox; Miller et al.) Overall growth in math achievement was significant, with some individuals making dramatic gains (e.g., moving through several years' curriculum in one year). Though anecdotal, positive results were also reported on other outcomes such as: Increased interest in mathematics, greater involvement in math competitions or extracurricular activities, improved self confidence about the study of math, and more positive attitudes toward school. The only negatives encountered were for some students involved in the CTY, in which the instruction occurred at Johns Hopkins University (JHU) (i.e., outside the students' normal school setting or curriculum). Some students were later forced by local school policies to repeat work they had already mastered at JHU and, on occasion, reported harassment from other students or local district professional staff on their return to school.

Ablard and Tissot (1998) have recently documented that academically gifted students (grades 2 through 6) can demonstrate abstract reasoning abilities sufficient to accommodate the study of algebra several years earlier than the typical chronological age of formal operational reasoning. They also report that there is no fixed chronological age at which this shift in thinking will occur among the gifted. While there was no reported measure of affective variables in their study, the researchers point out the potential for loss of motivation and development of poor study habits and attitudes toward learning if these students are not provided appropriate curriculum options before grade eight (the typical first year of accelerated access to the secondary math sequence). One can only guess what the child's attitude might be in situations such as those encountered by the Hopkins CTY participants.

Though not strictly based on the SMPY model, similar efforts have been made to accommodate mathematical talent for children at the kindergarten through second grade level (Robinson, Abbott, Berninger, and Busse, 1996; Robinson, Abbott, Berninger, Busse and Mukhopadhyay, 1997; Waxman et al., 1996). A sample of students was selected by locating 310 students in the greater Seattle area who scored above the 98%ile on one of the arithmetic subtests of the WPPSI-R, WISC-III, or K-ABC. A biweekly Saturday Math Club was provided to a random half of the group over a two year period. (The control half was also provided with a Saturday Club, but activities were more recreational than instructional.) Results showed that for both groups, the talent was maintained--was robust-- over the two years, but targeted enrichment enhanced the magnitude of the gains made by the children in the experimental section. [2] No results of affective or attitudinal factors were reported, though the 1997 study did note informal reports of some experimental group students who spontaneously brought the Saturday Club activities to their classrooms and enthusiastically shared them with classmates, suggesting positive attitudes on the part of the participating students.

Studies of Gender and Psycho-social Variables

In trying to find information about psycho-social variables affecting mathematical talent development, one is quickly drawn into the continuing, spirited and richly informative literature about gender differences in mathematics (Fennema and Hart, 1993). This seems to be the result of a convergence of societal trends in the early 70's (Arnold, Noble and Subotnik, 1996). Early SMPY data both stimulated great interest in precocious development and math talent, and revealed significant gender discrepancies in results on the SAT-M, favoring boys at highest levels by as much as 17 to 1 (Benbow and Stanley, 1980, 1981; Fox, 1976). This was also the era of intense societal awareness of feminist issues. In all arenas--political, entertainment, educational, philosophical, economic-- feminist perspectives were inescapable, and in psychology, the work of Carol Gilligan (1977, 1982) and her colleagues (c.f., Belenky, Clinchy, Goldberger, and Tarule, 1986) established a new paradigm for many psychologists and lively debate in the field which continues today. Maccoby and Jacklin's now-classic *The Psychology of Sex Differences* was published in 1974, and with regard to mathematics, seemed to establish that males of all ages and ability levels were somewhat more capable at mathematics than females of comparable general ability. In this milieu , the findings from SMPY were quickly seen as a failure-to-survive-a-critical-filter issue (Sells, 1973). (Sells sees excellence at mathematics as a "gatekeeper" quality for access to more lucrative and/or powerful adult roles. Seen in this light, the lower proportions of females attaining high scores are a systemic limitation to gender equity in the culture.) As Reis and Callahan pointed out in 1989, "More attention has been given to the issues relating to the potential and achievements of gifted females in the last five years than in the previous four or five decades" (p. 99).

Seeking factors to explain the gender difference in math achievement for the population at large, researchers in the last two to three decades have generated a substantial literature on gender discrepancies in math aptitude and attitudinal factors. Thus, most of what is known about attitude toward math is viewed through a lens of gender differences. A review of this body of research, both for the general population and the gifted, is offered here.

Gender Discrepancies in Math Ability

One of the most persistent problematic findings of SMPY is a gender discrepancy for extreme mathematics ability--a discrepancy which increased (to approximately .5 s.d.) as the ability in mathematics increased. The discrepancy is also reflected in later career paths where "above 700" males are disproportionately represented in the physical sciences (Benbow, 1988; Benbow and Lubinski, 1997; Benbow, Lubinski, and Suchy, 1996; Hyde, Fennema, and Lamon, 1990; Kerr, 1997; Lubinski, Benbow and Sanders, 1993). These authors refer to a surfeit of studies reported between the late sixties and mid-eighties which documented this difference in both the general population as well as in the students of high ability. "Sex differences in SAT-M scores among young adolescents are not temporary trends. They have been stable even in times of great change in attitudes toward women" (Benbow, 1988, p. 172).

Several authors further review and summarize theories from the extensive literature about why the larger gender differences at the high ability levels might exist. Issues of socialization are frequently mentioned, including differences in: degree of focus in interests and values; course taking choices prior to the test; perceived attractiveness of competing vocational and lifestyle options (Benbow, 1988; Kerr, 1997; Lubinski et al., 1993); confidence levels for females; perceptions of mathematics as a male discipline (Benbow, 1988); and sampling bias due to self-selection in SMPY studies (Becker and Hedges, 1988; Kerr, 1997). (Benbow also reviews the somewhat more limited research which, for the highly mathematically talented, seems to find little support for socialization issues as an explanation.) Biological explanations are also offered, including differences in neurological structures, chromosonal patterns and/or pubertal hormones (Kerr, 1997); differences in mathematical sub-skills possibly related to brain functioning, particularly spatial visualization and mathematical reasoning; and differences in physiological factors which also relate to things like allergic reactions and left-handedness (Benbow, 1988). Statistical issues (e.g., limitation of range, use of standard scores rather than effect sizes, and variance discrepancies which are greater for boys than for girls) are addressed as artifacts of the analysis which may enhance results favoring males (Becker and Hedges, 1988; Lubinski et al., 1993). Last, gender bias in the construction and item content of the tests (Lubinski et al., 1993) is yet another possible explanation.

Focusing specifically on gifted females, Callahan (1991), Reis and Callahan (1989), and Reis (1998) summarized research which not only documents discrepant math test results favoring males, but also the underrepresentation of women in higher-level career and professional ranks, many of which are math-dependent. Despite over two decades of considerable research attention to the issue, only modest progress has been made. For example, females are taking more courses in advanced math and science, but at a lesser rate than boys --who have also increased in science and math course taking over the same period. In addition, Reis and Callahan cite research which shows that females have significantly lower self confidence in math, a difference which increases from elementary to high school. They find psycho-social issues as the most probable causes, including: Parental influences on female's perceptions of ability and expectations for success; lowered self-perceptions after receiving lower (than expected) high school SAT or ACT scores, resulting in lowered aspirations; continued stereotyping of sex roles in society, particularly with regard to the perception of math-as-a-male domain; and effects of classroom grouping and teaching strategies which inadvertently favor males. They especially cite the evolving practice of using cooperative learning groups, where it was found that females were often ignored or relegated to "secretarial" roles by males in the groups, and the nearly universal predominance in public schools of mixed-gender classrooms in which more-assertive-by-nature boys may dominate.

Though none of the theories explaining the discrepancies offered by all these researchers has been completely discredited, none has proven sufficient by itself to explain the difference in math performance. In peer commentaries to Benbow's (1988) article which summarized gender differences in mathematical reasoning, Jensen (1988) reported similar discrepancies in some previously unpublished data (i.e., boys to girls ratio of 4:1 at the 95%ile), but he found even greater discrepancies in his data between different ethnic and SES groups. Rosenthal (1988) suggested that, with a .5 s.d. discrepancy (between males and females at the extreme high end of the achievement curve) would be equivalent to a Pearson r of .24. Thus he argued that effect sizes of any of the proposed causative variables would have to be large, indeed, to singularly account for the difference. Sternberg (1988) also questioned the practical importance of the effect. The critical threshold of math ability one needs to accomplish professional goals is probably well below the the 700 SAT-M-before-age-13 score Benbow has examined. Indeed, variables other than extreme math reasoning ability may impact eventual success. At least some of these variables may be considered issues of values as much as of science, but are worthy of mention in a milieu that seeks gender equity in every arena.

Two meta-analyses on gender discrepancies in math achievement for the population at large may partially answer this question of how much of a difference between genders really makes a difference. Hyde (1981) showed that gender differences (whatever their cause) account for less than 4 percent of the variance in math achievement for the general population, and Hyde, Fennema and Lamon (1990), while also finding increasing gender differences in achievement as the samples became more highly selected, found the overall differences in mathematics performance to be small. Kerr (1997) concludes her summation of the research with the following:

"...even the most august and rigorous math-related jobs in the world--for example, theoretical mathematician, astrophysicist, cosmologist--do not necessarily require the most extraordinary mathematical reasoning powers in the land. The vast majority of mathematically gifted girls, certainly all those who qualify by the talent search criteria, have the intellectual capacity for any math-related position existing today if to their intellectual ability they add the *training, confidence, expectations, attitudes and personality characteristics* needed to explore the concept of number or the beginnings of the universe" (p. 487) (Italics added).

The issue is far from settled, though most experts appear to have adopted an explanation which blends the various possible causes of gender discrepancy. Kerr (1997) suggests that at least with regard to students of high mathematics ability, the cumulative effects inherent in the socialization theories which attribute differences to affective and attitudinal variables seem the most favored.

The research on these affective and attitudinal variables with regard to high ability mathematics will be examined next. Even if statistical discrepancies in aptitude or achievement are of little real-world importance, the attitudes and psycho-social factors which may lead students of either gender to opt out of higher level math courses or of math-dependent careers have significance to many. They are important to the individual, who risks lost academic and career opportunities and income, and to the culture, which stands to lose the significant contributions of a major portion of its most able thinkers.

Affective and Attitudinal Variables

Eccles (1985), drawing on her substantial contributions to the field of gender-related issues, described a multi-faceted path model, the Model of Academic Choice (MAC), to explain achievement-related decisions. While the model was developed for persons of all abilities, she reasons that its application to the highly gifted is especially important, in-as-much as the likelihood that psycho-social factors, not aptitude, are the salient variables contributing to lowered involvement in math by high ability girls (p. 263). In brief, Eccles defines four clusters of factors as contributing to achievement-related decisions: expectations for success; choice-making as a self-actualizing activity; the mediators (e.g., parents, previous experiences, cultural sex-role values, etc.) of the perceived field of options; and perceived opportunity costs involved with any given decision. These interact in complex ways, leading to achievement behaviors which she identifies as persistence, choice, and performance. She posits that each of the contributing factors will be different for high ability girls than they are for boys, resulting in the differential pattern of math course taking.

Another, less complex, path model is proposed by Fennema and Peterson (1985), the Autonomous Learning Behavior Model (ALB). In their model, sex related differences in high-cognitive level skills (e.g., complex problem solving) are a direct result of autonomous learning behaviors (ALB's)-- choosing to engage in high-level tasks and persistence in working independently until success is experienced, which strengthens the internal belief system and engenders more success. ALB's are directly influenced by both the internal belief system (confidence, task usefulness, sex-role congruency, and attributional style) and by external and societal influences, which also influences the internal belief system. Though appearing at different points in these two path-driven models, there are several common elements, illustrated here:

<u>Eccles: MAC</u>	<u>Fennema-Peterson: ALB</u>
Child's task specific beliefs	Confidence
Cultural stereotypes of subject	Sex role congruency
Choice and persistence	Choosing to work, with persistence
Utility value	Usefulness

It seems plausible, then, to speculate that these variables may be particularly salient in defining the differential outcomes of math-related decisions. These variables function in the internal belief systems, but it is important to recognize that they are--as most beliefs would be--heavily influenced by beliefs held by others.

One of the most widely used instruments to measure the attitudinal variables is the Fennema-Sherman Math Attitude Scales (1976), developed to measure attitudes and beliefs which might contribute both to choices about math course taking and to success in math. There are nine scales in the original instrument: Confidence in learning mathematics, mathematics anxiety, usefulness of mathematics, mathematics as a male domain, attitude toward success in mathematics, effectance motivation in mathematics, mother's attitude, father's attitude, and teacher's attitude. (These last three are as perceived and reported by the student.)

Hyde, Fennema, Ryan, Frost and Hopp (1990) conducted a meta-analysis of 156 groups from studies between 1967 and 1988 which measured gender differences in math attitudes and affect. Seventy-three of the samples were middle school or high school age students; nine were students age 10 or below, and the rest were post-secondary school. Most of the samples were non-selective or moderately selective with regard to ability, though three were selected for math precocity in middle and high school. The significant finding was little effect size difference in attitudes toward math, with two exceptions. The first is that males stereotype math as a masculine domain considerably more than females do (overall study effect size = .90 favoring males), with peak values occurring during the high school years. The second is that males report more favorable attitudes on the part of adults toward their mathematics performance than girls, also with a peak effect size for the high school years. The report notes that this trend showed a decrease between the studies reported in the 1970's, and those in the 1980's, though it is not clear if that was due to a greater support perceived by females, or less support perceived by males.

Hyde et al. (1990) found that for the math-selective groups (originally reported in 1982), similar results obtained, though effect size on the math as a male domain measure is much larger: For the two high school studies, the average effect size was 1.95 whereas the single middle school study effect size was 1.60. This suggests that, among the most able students of mathematics, boys stereotype math as a male domain much more strongly than girls, and much more strongly than is the case in the general population.

On the variable of attitude toward math success, Hyde et al. (1990) reported the general population showed only a small effect size for these two age groups (middle school effect was .13 and high school was .06, both favoring boys). However, the middle school math selective study showed an effect size of .55 favoring boys. The younger high school math selective study found an even larger effect size of .79 favoring boys, while the older high school study effect size was .44 favoring girls (yielding a high school average effect size of .18 favoring boys). This suggests that high math ability boys have a much stronger belief, especially during middle school and early high school, that success at math is a socially desirable outcome. However, by later high school, the relatively greater strength of the boys' belief is attenuated as girls begin to also show a stronger belief in this variable, as well. The results reported here must be applied with caution, however, as the n was small (n=100 for all three age level samples combined), and the cited study at best reflected attitudes of two decades ago on an issue which has continued to stimulate significant discussion, if not change, during the intervening time.

As recently as 1994, Fennema and Hart found gender differences similar to those reported 1974, including differences in personal beliefs about mathematics. Reis, Callahan and Goldsmith (1996), in a qualitative study of highly able middle-school boys and girls, also found attitudes similar to those of earlier decades.

Affective and Attitudinal Variables: Elementary Years

As with the SMPY and other data on math ability, in the many studies summarized above very few were of children below the age of 12. A new wave of research into these variables for children during the elementary years, and as they move into junior high, has appeared since the studies cited above (Eccles, Wigfield, Harold, and Blumenfield, 1993; Frome and Eccles, 1998; Wigfield and Eccles, 1992, 1994; Wigfield, Eccles et al., 1997). The studies deal with the general population covering several areas of competence. Summarizing from them, several conclusions can be drawn from these studies with regard to beliefs about math in younger children:

1. Children's competence beliefs about math appear to be most closely related to their mother's perceptions of their math ability--i.e., their mother's interpretation of external information like grades and teacher feedback influenced children's self-perceptions more than their actual experiences in math.

2. The mother's beliefs, more so than the father's, tend to be stereotypical about mathematics, which in turn influences the child's belief due to the mother's stronger mediational factor.

3. Competence beliefs are somewhat higher for both genders in early elementary school than in the higher elementary grades when they become somewhat more realistic, though boys at all grade levels had more positive competence beliefs for mathematics than did girls. The junior high years brought significant decreases in competence beliefs for both genders.

4. Children's belief in the usefulness and importance of math was high and relatively stable throughout elementary school for both genders, decreasing only modestly

from first through sixth grade. Significant declines in the beliefs of both genders occurred from that point through junior high, however.

5. The valuing of tasks, for both boys and girls, has a direct relationship to their competency beliefs about them. This confirms earlier theories.

6. The valuing of math was relatively stable until fourth grade, when a significant decrease began that continued through junior high for both boys and girls.

Gender stereotypical patterns for math emerge around sixth grade. These first significant changes, labeled *gender role intensification* (Hill and Lynch, 1983), mark the point when the two genders begin to diverge in both self-perceptions and actual achievement. Also around this time, mothers, in particular, begin to somewhat underestimate their daughters' aptitude, attributing their daughter's math success more to effort than to ability. This dynamic reverses for mothers when considering sons. Fathers do not appear to differ as much in their views, but seem to play a critical role in defining a realistic view of math ability for children of both genders. These conclusions can provide a backdrop for consideration of similar information about children with high ability in math as they move through elementary school and into middle school or junior high. At this time, very limited data are available on this issue.

In this author's search, only four studies were located which examined any psycho-social variables in relation to high math ability in children under age 12. Ewers and Wood (1993) examined both math self-efficacy and prediction accuracy in 38 gifted and 38 average ability fifth graders, evenly divided by gender. The gifted group was selected by the local district's criteria, which were not specified. The Stanford Achievement Test math calculation subtest mean for the gifted group was the 97.2 %ile. As was hypothesized by the researchers, the gifted children showed higher self-efficacy expectations than the average students. Even though no gender-related difference in performance was found for either the gifted groups or the average groups, the boys at both levels had marginally higher self-efficacy than the girls. However, the gifted girls were the most accurate (of all four groups) in predicting actual performance--i.e., the gifted boys somewhat over predicted the accuracy of their performance, while both boys and girls of average ability, though having lower self-efficacy predictions, made larger over-estimates of their performance than did the gifted boys. Ewers and Wood proposed that a tendency to predict failure on problems which are subsequently solved correctly may lead gifted girls to avoid potentially rewarding math-related endeavors, and that the overly optimistic predictions of success in the average ability students may lead to disillusionment and progressive feelings of failure. The study was exploratory and limited in scope and caution should be taken in generalizing its results.

Essentially the same results, however, were obtained by Junge and Dretzke (1995) in a study of self-efficacy among mathematically gifted high-school students. Taken together, the two studies suggest that there may be a persistent gender difference in the perception of self-efficacy for mathematics, which is present as early as fifth grade, and which may impact negatively on girls with regard to many math-related decisions made during the middle and high school years.

In the second study, Hughes (1985) examined the relationship between the attitudes of gifted girls and boys (again, based on the school's criteria, which were not specified in the study) toward traditionally defined feminine roles, math achievement, and attitudes toward math related careers. No significant difference in attitude was found between the 76 boys and 56 girls in the Catholic fourth through seventh grade sample.

Cramer (1989) conducted a qualitative study, using in-depth interviews with four fourth grade students (two boys and two girls) who were members of an advanced math class in a public school gifted (I.Q. > 140) program. Her main finding was that stereotypical thinking dominated the boys' responses to nearly every question, even in the face of evidence to the contrary (e.g., one of the boy's mothers was a vice president of a local bank.) The girls, however, believed that females are smarter than boys, but fail to achieve as highly due to lower self confidence, shyness, and lower physical strength. The girls (but not the boys) were also clearly aware of the effects of sexism on women's achievement in the workplace, and reported resentment of their male classmates' disparaging attitudes. She concluded that over time the boys' attitudes may account in part for the girls' lowered achievement and interest in mathematics.

Swiatek and Lupkowski-Shoplik (2000) recently contributed to the literature on attitudes of high-ability elementary aged students, addressing the possible relationship between liking for academic subjects, and tested ability in those subjects. The sample of 2089 third through sixth grade students were all first-time participants in an elementary level SMPY-model testing program, selected by scoring above the 95%ile on one or more subtests of their grade-level standardized tests--that is, they were not selected only for math ability. The students were asked to complete a brief survey at the time they took the out-of-level EXPLORE test. A prompt stem, "How do you feel about...", was used for ten different school subjects, plus a "school in general" question. Students responded to a four-level Likert scale. Between the third grade and sixth grade cohorts, there was a statistically significant ($p < .005$) difference in liking for math, which occurred incrementally across all four grades. The difference in liking for math was comparable overall to the difference in liking for writing, art/music, and reading, while even greater differences were found for computers and for school in general. Thus, while math is less well liked by the high ability students at the sixth grade than at the third grade, it appears to be part of a general decline in liking of a number of features of school. The correlations between liking for school subject and EXPLORE scaled scores proved to be negligible to small. The researchers note that correlations might be apparent if the sample were less selective. They also note that this lack of correlation is parallel to findings for high ability high-school aged students, and speculate that "participants reports of liking for various subjects reflect more than just their tested ability level in those areas" (p. 373).

An overview of the above findings is provided in Table 1.

Responding to High Ability

As noted in the SMPY and SMPY-related studies above, targeted interventions involving different curricular options, particularly acceleration, were highly successful with the high ability students at all ages (Benbow and Lubinski, 1997; Benbow et al., 1996; Miller et al., 1995; Robinson et al., 1997; Stanley, 1996). This is not surprising, confirming the traditional opinion of experts (cf., Daniels and Cox, 1988; Feldhusen, Proctor and Black, 1986; Terman and Oden, 1947; Van Tassel-Baska, 1981; Webb, 1983) and empirical studies of the effects of acceleration (Delcourt, Lloyd, Cornell and Goldberg, 1994; Kulik and Kulik, 1984; Rogers, 1991).

Issues With Acceleration

The use of the term *acceleration* can create difficulties: is the subject matter being delivered faster, or are students being provided administrative options to allow access to course work at a younger-than-expected age? Southern and Jones (1991), summarizing the extensive literature on accelerative options, define fifteen different forms of acceleration, each with a different balance of these two factors. Two issues are problematic in those options which place students in more advanced course work, in contrast to adjusting the pace of instruction once the material is accessed. [3]

The first is related to concerns around negative psycho-social outcomes for

Table 1

Studies of Gender Discrepancy on Math Achievement and Attitude Measures

An Overview of Findings Comparing Gifted and General Populations at Three Age Levels

	Math Achievement	Math Attitude
Elementary Age		
General	Small difference favoring girls on computational tasks; no significant difference overall.	Nonsignificant differences.
Gifted	Seems to favor boys (data limited to one study).	Higher in grade 3 than 6; May favor boys (data limited to three studies).
Middle School Age		
General	Small difference favoring boys, especially for problem solving activities.	Small differences: Boys see math as a male domain more than girls; girls lose confidence more than boys; mothers attribute girls' achievement to effort, boys' to ability.
Gifted	Favors boys, by as much as 17 to 1 at the highest level (4:1 for the top 5%ile).	Boys see math as male domain and math success as more desirable (limited, but consistent data).
High School Age		
General	Nonsignificant difference, favoring boys.	Small differences favoring boys.
Gifted	Differences found at early adolescence increase; results confounded by course taking which favors boys.	Strong differences favoring boys in perceiving math as a male domain, and of support by adults for math

students who are pushed into situations they are not developmentally ready to handle. Southern and Jones (1991) define several areas which have emerged in the literature: Potential difficulties with the advanced material due to gaps in prerequisite knowledge or inability managing increased expectations; social adjustment difficulties due to removal from chronological peer groups; reduced extracurricular opportunities, especially in athletics; and emotional maladjustment due to adverse effects in academic and social arenas. Southern and Jones point out that a large body of research shows these adverse effects are more feared than real. However, they also note that the data in most research is aggregated, while the negative results are of primary importance to individuals, and suggest the question of whom to accelerate is much more important than whether acceleration is good.

Stanley (1979) addressed this particular point indirectly, in his case-study report on the first 14 of his now-classic "above 700 before age 13" students who began, at junior-high age, to take college-level math classes at Johns Hopkins University. He reports that 13 of the students did extremely well, not only succeeding beyond the level of their college student classmates, but also seeming to thrive in the more serious milieu of the college campus. The 14th , he notes, initially stumbled due to choosing to take an overload of particularly challenging college-level classes (against SMPY staff advice), and then refusing to study (based, one presumes, on the lackadaisical habits he had developed in his regular school). This student, however, eventually adjusted his expectations, improved his study habits, and by the end of the second semester, was as successful as the other 13. Subsequent research generated by SMPY, (cf. Benbow, Lubinski and Suchy, 1996; Charlton, Marolf, and Stanley, 1994; Lubinski, Webb, Morelock, and Benbow (2001); Swiatek and Benbow, 1991) as well as several meta-analyses which are detailed in the following section have substantiated Stanley's early reports.

The second issue is much more salient to the current study. Most schools implement acceleration through administrative adjustments in the school setting--they move students into next year's book and/or class (Southern, Jones and Stanley, 1993). This typically results in the *de-facto* creation of an ability group. The group may be a separate class (e.g., an honors class), or may exist as a small cluster of students within a larger class. Either way, these students have been selected based on some assessment of ability or advanced achievement and, in a quasi-public milieu, are being taught and can be observed achieving at a level which is higher than their age peers. In many quarters, the value of acceleration *per se* is not debated, but the concomitant ability grouping (whether for remediation or acceleration) is a topic which generates strongly held opinions. Intense debate about this issue has resurfaced in educational circles in the last decade and a half.

Issues with Ability Grouping

Though an issue which has a history of ebb and flow throughout this century, the current debate on grouping seems to have been sparked by the seminal book, *Keeping Track: How Schools Structure Inequality* (Oakes, 1985) and a best-evidence synthesis reported by Slavin (1987). According to these reviews, not only is ability grouping an ineffective educational practice, but it creates serious risk to the self-esteem and ambition of many learners, is driven by racist attitudes, and is an affront to democratic values. The belief that the democratic value of equity demands the abandonment of ability groups has hindered the practice of acceleration--for which grouping is a common, though not universal, delivery model-- despite strong research evidence supporting its effectiveness (c.f., Benbow and Lubinski, 1997).

Policy makers and school officials all over the country have reconfigured schools over the last decade to conform to these perspectives (Kulik and Kulik, 1991, 1997). Reflecting beliefs stemming from the civil rights movement, those who hold equity as the most salient value in educational practice offer solutions in which teachers differentiate within the classroom to meet the instructional demands brought to the classroom by students of widely varying abilities and skill levels (c.f., Reis, Burns and Renzulli, 1991; Renzulli, 1994; Tomlinson, 1995, 1999). (The present discussion focuses only on the gifted, but parallel strands of the debate are readily found in relation to other special-needs students, e.g., the learning disabled. The title of Kauffman and Hallahan's 1995 book, *The Illusion of Full Inclusion : A Comprehensive Critique of a Current Special Education Bandwagon*, gives a sense of the debate as it is playing out in other arenas.) It is believed that not grouping students by ability will prevent damage to self-esteem and lowered achievement which is perceived to accrue to those placed in the low groups, and will equalize educational opportunities and outcomes by bringing engaging, conceptually rich, high-level instruction into every classroom, rather than (it is asserted) reserving such instruction only for the gifted (Oakes, 1985; Slavin, 1987; Robinson et al., 1996).

Recent research has begun to cast serious doubt on these conclusions, however. Kulik and Kulik (1991, 1997), in meta-analytic reviews of research on ability grouping's effects on all learners, found significant academic achievement for the gifted, modest gains for the middle, and no difference academically for students with low ability. The findings on affective variables in both these meta-analyses were less clear, primarily because they were reported far less often. However, the trend appeared to be the exact opposite of what Oakes (1985) and Slavin (1987) claimed--i.e., Kulik and Kulik found that "[when ability grouping is used]...quick learners lose a little of their self-assurance, and slower learners gain some badly needed self-confidence " (1997, p. 240).

Fuligni, Eccles and Barber (1995) report a longitudinal study comparing tenth grade outcomes of seventh graders, either grouped and non-grouped by math ability. The grouping had positive immediate (i.e., in the same same year) effects on both the academic variables and the psycho-social variables. By the tenth grade, the seventh grade grouped-for-math students, in comparison to the seventh grade non-grouped students of the same ability, had greater positive outcomes on all variables for both the medium and high groups. The grouped-for-math low students showed either positive or null effects for all variables except achievement, which showed a statistically insignificant negative effect. The authors note that all effect sizes were small, which they saw as predictable for a distal outcomes study design.

Rogers (1995), in a meta-analysis of a variety of accelerative options (many of which involve ability grouping), reported positive effect sizes for academic variables and positive or null effect sizes for affective variables for each of the options (see Table 2). In a review of grouping practices, many of which are accelerative in nature, Rogers (1991) notes "It is unlikely that grouping itself causes academic gains; rather, what goes on in the group does" (p. 2).

Table 2

Effect Size for Accelerative Options

Likely to Involve Grouping at the Elementary Level[a]

	Academic	Social	Self Concept
General Impact Across All Options			
Grades K-2	.64	.20	.16
Grades 3-6	.59	.20	.16
Impact By Accelerative Option			
Nongraded Classroom	.38	.02	.11
Compressing Curriculum	.45	---	---
Subject Acceleration	.49	---	-.16
Enrichment Pull-out/ Subject Extension	.65	---	.11
Cross-Grade Grouping Reading, Math	.45	---	---
Separate Classes for Gifted	.33	---	-.14
Regrouping for specific instruction	.34	---	-.06
Cluster Grouping	.66	---	---

[a] An effect size of .33 is considered statistically significant *(sic)*

(From Rogers, 1991, 1995)

What goes on in the group may involve factors which are not specifically academic, however. Hunt (1994, 1996/97) reported results which seemed to suggest that the group composition, alone, exerts sufficient influence to create a positive difference for gifted learners. She found statistically significant positive effect sizes in both math achievement and attitude toward math for the ability-grouped gifted sixth graders when compared to gifted students taught the same curriculum, using the same teaching techniques and strategies, in mixed ability groups. Yet she found no statistical significance between the two types of grouping arrangements in achievement for medium and low ability students learning the same curriculum. Thus, the positive outcome may be more attributable to the grouping than to the curriculum, at least for the gifted. This would support the instructional value of the interaction of both spontaneous teacher responses and students with similar abilities--a dynamic unlikely to accrue to students in mixed ability groupings.

Smith, Jussim, Eccles, VanNoy, Madon and Palumbo (1998) studied the impact of grouping on self-fulfilling prophecies among sixth graders who were grouped by ability, both within a mixed ability class, and in ability-grouped classes. Their primary findings were that (1) teacher perceptions predicted student achievement primarily because those perceptions were accurate; (2) marks for students in mixed ability classes were skewed in a direction consistent with this accurate perception; and (3) self-fulfilling prophecies were strongest when students were grouped within classes, though they were typically very small. They conclude that, in the complex social context of classrooms, teachers and children alike make social judgments about each other based on a wide variety of information, and that for the most part, these judgments are surprisingly accurate and not notably skewed by the grouping labels.

Page and Keith (1996), using a path model, analyzed data from the 1980 High School and Beyond data base, a massive study which included information about student ability as well as post high school outcome variables. Their conclusion is that grouping by ability (not specified by any particular subject) has positive effects on high-ability students' life achievements, with a substantively stronger effect for high ability minority youth, and no negative effects on the achievements of low ability groups. Additionally, they found no negative effect on educational aspirations, self-concept or locus of control related to students' ability grouping, for any group studied. There was a small positive effect on educational aspirations for high ability students in general, and a moderate positive effect for high-ability Hispanics.

Several conclusions can be drawn from this emerging body of research. First, it is erroneous to assume that not grouping students is the way to prevent negative social labels or negative affective or psycho-social outcomes. Altering the level and pace of mathematics instruction in ability grouped settings has clear positive effects on academic outcomes for middle and high ability students, and null or, on occasion, statistically insignificant negative outcomes for those of low ability. Attitudes, when they were studied, were strongly positive for the gifted placed in like-ability groups, and--significantly--were not diminished for other students. These conclusions appear to refute the contentions of Oakes (1985) and Slavin (1987) and support the value of ability grouping, at least for mathematics instruction. Furthermore, considering Oakes' and Slavin's strong belief in educational equity, it ironically appears that ability grouping is especially effective for high ability minority students, at least at the high school level (Page and Keith , 1996).

Loveless (1999a, 1999b) discusses important questions related to the secondary effects of failing to group for ability: Issues of school climate, social equity, and general student self-esteem. He points out that society sends a strong negative message about the value of achievement when schools are organized to purposely hide evidence of high achievement. He further posits that it is the brightest students among the minority and disadvantaged who stand to lose the most when schools do not provide access to strong, achievement-oriented curricula. The achievement and drop-out gap data on different racial groups shows a decrease during the fifteen years preceding the Oakes (1985) and Slavin (1987) reviews, when ability grouping was fairly broadly used, but an increase for the fifteen subsequent years when ability grouping has been largely abandoned. He notes there is even greater risk to minority or disadvantaged students, because parents of means will abandon public schools for more challenging options--*bright flight* (p. 30, 1999b)--when challenge is deleted from the curriculum. Failure to acknowledge and provide for individual differences among students further diminishes general self-esteem of students, who experience homogeneous grouping as a "herd mentality" on the part of the institution.

Other Related Research

Two substantive strands of research have recently evolved which, though tangential to the core questions of the current study, are briefly reviewed here due to their high visibility in both professional literature and public media, and their generally informative value to the current study's focus. The first of these are the many studies on gender equity in schools which developed from the seminal report sponsored by the American Association of University Women (AAUW). The second is the emerging literature generated by the Third International Mathematics and Science Study (TIMSS).

Gender Equity In Schools: The AAUW Report

The seminal report of the AAUW, *Shortchanging Girls, Shortchanging America* (1990) catalyzed public awareness of the many inherent and systemic disadvantages experienced by females in their K-12 schooling--an issue which had been developing in the professional literature for two decades. Of particular concern was the finding that girls, around puberty, experienced a much more precipitous decline in self image and career expectations than did boys of the same age; at exit from high school, they were significantly lower than their male peers. The report also found a circular relationship between self-esteem and enjoyment of math and science, among both boys and girls--greater self esteem students reported greater liking of math and science, and those who like math and science had higher self esteem scores. Even so, the career aspirations of girls in math and science areas were much lower than boys, and declined to a much greater degree as they moved through secondary education. Relationships with teachers were the strongest predictor of girls' academic self confidence and career aspiration, while boys were influenced by their grades and potential to "do things" in a career.

Soon after the original report, the AAUW published a second report, *How Schools Shortchange Girls: A Study of the Major Findings on Girls and Education* (1992) which examined inequities in curriculum and instructional practices. Findings from these two publications were widely disseminated, enhanced by a number of additional books and papers targeted to the general public (cf., *Meeting at the Crossroads* (Brown and Gilligan, 1992); *School Girls: Young Women, Self-esteem, and the Confidence Gap* (Ornstein, 1994); *Reviving Ophelia: Saving the Selves of Adolescent Girls,* (Pipher, 1994); *Failing at Fairness: How Our Schools Cheat Girls* (Sadker and Sadker, 1994)).

There are many critiques of the premises of the AAUW reports. Though acknowledged as compelling, they are perceived by some to be "...a one sided, biased report by a group with an agenda--it's special pleading for a subset of the population. And it's a familiar problem: a group with its own agenda producing a report supposedly substantiating what it's been arguing" (p. 3) (Finn, quoted in Shields, 1992). Sewall, editor of *Social Studies Review*, focused on the emphasis the 1992 report placed on schools' social studies curriculum. He believes that the authors "are either ignorant of today's curriculum or --more alarming even---unaware of the hypothetical biases through which [they] view education" (Sewall, 1993, p. 94), and he criticizes the AAUW curriculum review as "based largely on outdated studies, many of them obscure and others politically transparent," noting that "curriculum ...should never be confused with a program of therapy. ...[Curriculum] is designed to instruct, not proffer remedies for human and social failings" (p. 96). Other reviews are somewhat less acerbic, considering the report important but perhaps off-point in describing problems which accrue to girls at school, in an era when there is widespread concern for the decline of student achievement across the board. It is limited for suggesting only curriculum revisions when those must be undergirded by wholesale educational reform (Giarelli, 1994; Shields, 1992).

While informative about an important developmental issue, this body of literature does not attempt to address the specific issues of math attitudes for gifted students, the focus of the current study. As noted in the preceding literature review, much of the research investigating attitudes toward math was conducted along gender lines due to the evidence of gender discrepancies in math achievement and the concerns for gender equity which developed in the 1970's. This literature (reviewed above) seemed to find evidence of a drop-off in attitudes regarding mathematics (as well as school in general), and where genders were disaggregated, there appeared to be greater decline for girls than for boys--findings congruent with the AAUW reports. The current study, therefore, can anticipate a decline in attitude toward math. The degree of decline for the highly precocious students in relation to their chronological peers is unknown.

U.S. Mathematics Instruction Compared To Other Countries

The other research strand with relevance to the current study is the Third International Math and Science Study (TIMSS), an extremely comprehensive international study of mathematics and science achievement, curriculum, social factors, and instructional practices. TIMSS has garnered significant attention in the popular media as well as generated numerous publications in the professional literature since its completion in the mid-1990's. (TIMSS data and study-generated publications can be accessed at http://ustimss.msu.edu/). With regard to math, TIMSS found that the United States is slightly above the middle (of the 50 countries) in achievement at the fourth grade level, declining to the mean by the eighth grade, and severely dropping in comparative achievement by the twelfth grades. At that level, the U. S. achievement was better than only two other countries--Cyprus and South Africa (Cogan and Schmidt, 1999). Though international comparisons become more problematic at the high school level due to wide variations in curriculum and societal standards for school completion, even the top 14% of American students--those enrolled in precalculus, calculus or higher courses--did not fare well in comparison to their academic peers in other nations (Cossey, 1999). The TIMSS researchers attribute the dismal results of U.S. students to the interaction of systemic factors: Unfocused curricula, characterized as "a mile wide and an inch deep"; unfocused textbooks, covering an excessively (in comparison to other countries) broad range of topics and overloaded with review and repetition; and teachers caught in a "breadth vs. depth" direct instruction approach because of the many topics they are expected to cover, despite often knowing and preferring alternate strategies (Schmidt, McKnight and Raizen, 1997).

Though the purpose of the TIMMS research is far broader than the focus of the current study, some of the findings may have some relevance to the current questions about the experiences and attitude of gifted math students. The 1997 report emanating from the TIMSS researchers, *A Splintered Vision* (Schmidt, McKnight and Raizen), finds little vision within American math education. The authors note that even those weak visions which do exist are diverse: Driven by varied social agendas, implemented in a system of distributed educational responsibility, and fueled by an assembly-line ideology resulting in "many small topics, frequent low demands, and interchangeable pieces of learning to be assembled later" (p. 8). In this milieu of diverse learning goals, the authors speculate "they [the goals] likely lower the academic performance of students who spend years in such a learning environment" (p. 2). While their concern is for all students, it is logical to speculate that this sort of fractured instructional milieu, when encountered by gifted students who are *de facto* already conceptualizing math well beyond their peers, would be particularly disheartening and likely to lead to a significant decline in motivation, interest and eventual achievement. Mathematically gifted students could hardly be expected to be captured by the content of a shallow, patchwork math curriculum which, by the eighth grade, is a full year behind that of many other high-achieving countries (Reys, Robinson, Sconiers and Mar, 1999). This seems to be supported by the discouraging results noted above for the top 14% cohort--a group which would include a high proportion of the nation's most mathematically gifted students.

This viewpoint, in combination with the grouping and acceleration studies presented above, begins to suggest another dimension with regard to mathematics learning which has not been studied to date. The failure of schools to provide appropriately for the brightest math students may well increase negative attitudes about either the subject matter or the importance of achievement which could, in turn, compound the impact of the lack of access to appropriately challenging curriculum. This would be an especially important dynamic in the pre-middle school years, when students' attitudes about mathematics are still in the formative stages. The opportunity cost to society--if it inadvertently teaches bright young children that working hard in math is not important, and that being very good or even better than other classmates is something to hide--is potentially great, indeed. The secondary school and eventual career trajectory of students who may absorb these "hidden messages" could be expected to be flat, and would probably resemble the low-math involved life choices found in the literature reviewed above. The current proposal is for a study to examine the possibility that failing to provide access to appropriate curricular challenge through acceleration of content for bright math students in the mid-elementary grades may cause a significant subsequent decline in favorable attitudes toward math and math achievement.

Summary and Critique

This review has addressed several disparate strands of research literature, to provide the context for the current proposal. Though not of primary interest in the current proposal, the research on gender differences in mathematics was reviewed because this has been the issue which has driven research on math attitudes for the last three decades. The impact of acceleration of gifted secondary math students has been studied through the SMPY longitudinal research. Because of the difficulty in sorting out the multiple forms of acceleration from both the practice and politics of ability grouping for elementary students, the issues for the gifted elementary students involved in like ability vs. mixed ability grouping were reviewed even though ability grouping, *per se*, is of only tangential interest to the current proposal. Lastly, two major extant research efforts, the AAUW gender equity studies and the TIMSS research, were briefly reviewed because of their broad dissemination and their potential informative value to the questions raised in the current research proposal.

The many reports which have developed from the SMPY studies seem to clearly show the effectiveness, for both academic and psycho-social outcomes, of acceleration in mathematics for the very brightest math students at the secondary level and beyond. Thirty years of research for students above age 12 is beginning to be supplemented with an emerging body of research for similar ability students before age 12. To date, the findings for the younger students appear to be similar for academic outcomes. Though less well documented empirically, it is none the less suggested that accelerative options also yield positive effects on the attitudes and motivation for the younger age students.

There are still major gaps in the information base. Table 3 summarizes the nine studies, listed in chronological order, with relevance to issues of math attitude among gifted elementary students. Two of the seminal reports on the SMPY longitudinal study

Table 3. Summary of Key Findings of Empirical Studies of Young Gifted Math Students

Authors	Type of Study	Statistical Qualities	Subjects and N	ID Process	Controls	Instruments
Benbow & Arjmand 1990 Benbow & Lubinski 1997	Longitudinal: 50 years from inception using periodic surveys and statistical analysis	1990: Descriptive and step-wise linear discriminant function, $p < .05$ 1997: Descriptive	Age 12 through mid-twenties, greater Baltimore, Iowa and national samples n=6000	6 cohort groups using SAT scores: 72-74: V ≥ 370 or M ≥ 390 76-79: top 1/3 of SMPY group 80-83: V ≥ 630 or M ≥ 700 87-90: V ≥ 430 or M ≥ 500 like 76-79 group 92: grad students in top-ranked engineering, math, or science programs Comparison groups 82: M ≥ 500; 83: C ≤ 540	Comparison groups	SAT scores at age 12 Surveys for 5-year cycle of follow-up studies
Hughes 1984	Research: Correlation between attitudes of gifted girls towards traditionally defined feminine roles, math achievement and math career aspirations, as compared to boys	Descriptive	Grades 4-7, New York City Catholic schools n=123	Gifted by local criteria, unspecified	Boys, matched for ability	Scott-Foresman Achievement Test in Math Pupil Perceptions Test Of Child and Adult Sex-Typed Roles At School and Home Researcher designed questionnaire on attitudes toward math careers
Daniels & Cox 1988	Report on Center for Talented Youth Program	Descriptive	Grades 2-6, greater Baltimore n=59	High scores on SCAT-Q		SCAT-Q for entry STEP to show mastery = 90% at exit
Cramer 1989	Qualitative in-depth interview		4th grade students Fairfax, VA n= 4 (2 m, 2 f)	Gifted program IQ > 140		Interview- 3 sessions
Ewers & Wood 1993	Research: Relationship of self-efficacy to both performance and prediction accuracy for math, for gifted and non-gifted students	Inferential, $p < .05$ Multivariate: 2x2 ANOVAs for gender x ability	5th grade, central New York state n=76	Non-gifted: Cognative Abilities Test < 111 Gifted: District criteria, unspecified Stanford Achievement Test-Math Calc. M = 90%ile		Research constructed 5-point Likert Certainty Scale Researcher constructed 20 item math task of mixed difficulty items, from easy to challenging
Hunt 1994 1995/6	Research: Effect of grouping on gifted students' attitude and achievement in math Quasi-experimental, pre-post test design.	Descriptive ANCOVA for pre-post mean gain difference ANOVA on grouping questionnaire	6th grade, Southwestern suburban n=208: 32 gifted, 106 average, 70 low	Gifted: CogAT-Math ≥ 90%ile (grp. A) Otis-Lennon > 120 Behavioral Rating Scales Torrence Test of Creative Thinking- Figural (grp. B) Average: CogAT- Math between 41- 89%ile Low: CogAT < 40%ile	Heterogenous and homogeneous groups matched for ability	TOMA Story Problems and Computation subtest, pre- and post TOMA attitude subtest, pre- and post Researcher constructed 10 item questionnaire on effects of grouping, 4 point Likert scale
Miller, Mills & Tangherlini 1995	Report on Model Math Program	Descriptive	Grades 2-6, central Pennsylvania n=456, over 6 years	IQ > 130 plus grade level standardized achievement ≥ 95%ile (Families could opt out)		IQ test Grade level standardized achievement test SCAT-Q out-of-level used for group placement: > 65%ile given rapid acceleration STEP subtest for diagnoses and post-treatment

Table 3 (cont'd)

Authors	Key Finding: Academic	Key Findings: Attitude	Key Findings: Other	Limitations
Benbow & Arjmand 1990 Benbow & Lubinski 1997	High score at 12 strongly predicts achievement at 22 Approximately 50% chose science/math careers	Students view SMPY interventions as positive, both in the short and long term Contact with intellectual peers valued, especially by females	High college achievement best predicted by (in order of strength): -pre-college curricula in math and science -family characteristics and educational support -attitudes toward math, and science aptitude High ability stimulated by special opportunities is strongest predictor of achievement at 22	Participation in talent search is self-selected Lifetime patterns likely to be influenced by societal changes, limiting generalizability Study does not address pre-adolescent years
Hughes 1984	Number of gifted girls declined after grade 5, but remained stable for boys	No significant correlations found between achievement and either attitude survey, for either boys or girls		Criteria for gifted not clearly indicated, may not be math focused Catholic school sample may limit generalizability Comparison to boys, rather than to non-gifted
Daniels & Cox 1988	54 achieved > 1.5 years growth in 48 hours of instruction, spread over 1 year	Student motivation and interest increased	Social/emotional development enhanced	Participation in talent search and courses is self-selected n is relatively small No follow-up on older grade-level outcomes Method of collecting attitude data and other findings not defined
Cramer 1989		Boys hold strongly stereotypical views Girls hold gender-neutral views and resent sexism by boys		Small n Not quantitative
Ewers & Wood 1993	No gender related differences in performance	Gifted had higher self-efficacy beliefs than average students Boys tended to over-estimate self-efficacy		Criteria for gifted not clearly indicated, may not be math focused n is relatively small, after disaggregation
Hunt 1994 1995/6	Homogeneously grouped gifted students completed more math activities, $p < .05$	Grouping patterns had little impact on change in attitude toward math, for all levels, $p > .09$	Gifted students preferred working alone and not in mixed ability groups Gifted students had strong and consistent written opinions about grouping, favoring homogeneous settings	Use of two gifted criteria confounds results, n's for each group, limits generalizability Relatively low criteria for gifted, compared to most other studies Ceiling effects on TOMA severely limited possible pre-post change for gifted groups 12 week duration may not be long enough to document attitude change Attrition rate (16%) is high for 12 week study
Miller, Mills & Tangherlini 1995	Initial rapid acceleration for > 65%ile's, leveling off as s's reached instructional level All > 65%ile's ready for algebra by 7^{th} grade, some by grade 5 or 6	Increased interest Improved self-confidence More positive attitude toward school	Increased participation in math clubs and competitions Higher mean scores on regional math tests Doubled number of math awards or honors	Participation in talent search and courses is self-selected Initial IQ screen not typical for SMPY models

Table 3 (cont'd)

Authors	Type of Study	Statistical Qualities	Subjects & N	ID Process	Controls	Instruments
Robinson et al. 1996 Robinson et al. 1997	Research: Longitudinal, experimental design to investigate developmental changes and gender differences in mathematically precocious young children	Variety of multivariate statistical tests applied to various phases of the research	Pre-K and K, greater Seattle n= 284	Parent recommendation and K-ABC Arithmetic subtest ≥ 98%ile, or WPPSI Arithmetic subtest ≥ 98%ile, or WISC-III Arithmetic subtest ≥ 98%ile	Saturday enrichment group, matched for ability	15 different standardized pre-post measures of verbal, quantitative and spatial abilities
Abblard & Tissot 1998	Research: Strength of abstract reasoning abilities in elementary academically talented students	Descriptive Inferential, $p > .05$ MANOVA	Grades 2-6, greater Baltimore area n= 150	Grade level standardized achievement test ≥ 97%ile SCAT-Q > 70%ile, two grades out-of-level		SCAT-Q Arlin Test of Formal Reasoning STEP
Swiatek & Lupkowski-Shoplik 2000	Research: Gender differences in the relationship between giftedness and academic areas	Correlations between subject areas MANOVA for gender and grade-level effects on liking for each subject	Grades 3-6, greater Pittsburgh area n= 2089	First time participants in C-MITES, an elementary student Talent Search; grade level achievement test ≥ 95%ile		Researcher constructed 11 item, 4 point Likert scale questionnaire assessing liking of school subjects

Note: Studies are listed chronologically, to show evolution of research findings.

Table 3 (cont'd)

Authors	Key Findings: Academic	Key Findings: Attitude	Key Findings: Other	Limitations
Robinson et al. 1996 Robinson et al. 1997	More boys than girls show math precocity, even as young as age 4 Math precocity is robust from preschool to 2nd grade Precocious math ability is correlated with both verbal and spacial ability in young children Intervention increased the growth in mathematical abilities, over controls	None studied: Informal reports of enthusiastic sharing of treatment activities with regular classmates in home school	Math precocity can be accurately identified as early as age 4	Parent self-selected subjects
Abblard & Tissot 1998	Academically talented students may be ready to study advanced math several years before chronological peers Wide variability among the > 97%ile group suggests no fixed age when children are ready for advanced math	No findings: Speculation that absence of rigor may result in negative attitudes, poor work habits, and underachievement		Addressed only academic question- no data on attitudes
Swiatek & Lupkowski-Shoplik 2000	No relationship between tested ability and liking for subject	Gender difference in liking for math was the least of all areas, and considered negligible Liking of math declined from grade 3 to 6, $p < .05$		Single dimension scale limits usefulness of data Limitation of range on ability measure may impact on null relationship between ability and liking

for the older age students are also included, to facilitate comparisons due to the scope of that study and its influence on the thinking of those which followed. This small body of studies of the younger students include a number of important limitations and biasing factors, some fairly severe.

For all the SMPY-model studies, including the longitudinal, the students (or their parents) were self-selecting to be involved in a talent search or advanced program. Thus, more positive outcome attitudes might be anticipated, due to family values about math which preexisted the opportunity to participate in the math testing or opportunity. In some cases, the degree of parental commitment--a likely (though not documented) correlate with the family's valuing of math education-- would have to be significant, e.g., to get children to "downtown" classes every Saturday for a year or more.

With a single exception the reports given on attitudes were not studied directly in any of the younger SMPY-model studies; Miller et al. (1995) conducted parent surveys which showed positive results, but for a heavily self-selected group. The technical features of Ablard and Tissot (1998) appear strong, but the study's focus was strictly cognitive, and offers only speculation on the topic of attitudes. Swiatek and Lupkowski-Shoplik (2000) designed a study which avoids most of these limitations, but the survey instrument was very general and of limited usefulness with regard to attitudes specifically for math, as results were based on only a single question for each subject. Cramer (1989) reported on math attitudes found in qualitative interviews, but for only four students.

The Ewers and Wood (1993), Hughes (1994) and Hunt (1994) studies have serious problems with defining the criteria for the gifted group, and also are limited by relatively small *n*s. In the SMPY longitudinal study (Benbow and Arjmand, 1990; Benbow and Lubinski, 1997), there are several cohorts of gifted defined by various scores on the SAT. While the reports note that those at the "lower" levels had relatively "weaker" outcomes, all subjects had very strong math ability to initially qualify for the talent search. The Ewers and Wood, Hughes, and Hunt studies are at best studying a much more modestly gifted group than that in the SMPY Longitudinal study, which may partially explain their weaker results. In addition, the results which were reported in Hughes' and Hunt's studies were confusing and indecisive. It is impossible to know if stronger results might have been shown with more stringent criteria or measures, or more powerful statistical methods. In a sparse field of research, two "mushy" studies and a third with "null results" can blur the emerging picture considerably.

Proposed Study

Rationale

Anderson and Bourke (2000) note that students possessing more positive affective characteristics are viewed as more likely to acquire more and higher cognitive (and other) objectives. They define a three-stage model: "affective entry characteristics" (what students possess when they arrive at their schools and classrooms); "affective characteristics in schools and classrooms" (the lens through which students perceive and react to the events and activities that take place at school); and the affective characteristics students take with them when they leave. These third-stage affective characteristics are of two types. "Affective outcomes" are those characteristics which are incorporated into school or classroom goals and "affective consequences" are unintended results of schooling.

Attitude is one of many affective characteristics, and is also only one of many real-world variables which could impact on student choices about math activities such as courses, clubs, and time invested in study. Meuller (1985) points out that attitude measures predict behavior patterns better than they predict isolated behaviors. Thus, though a strong, positive attitude might not be sufficient to predispose a student to a particular decision--say, participation in a specific math extra-curricular club--the positive attitude of the student, as well as those of the significant others in his or her environment, would seem to be a necessary prerequisite for mathematics oriented behaviors. A significant body of research examines attitude variables affecting mathematics learning for the general population of elementary age students. These studies address the impact on future math performance of attitudes at the "entry stage" (Anderson and Bourke, 2000). Most of the secondary studies measure "outcome or consequence" attitudes which may result from math experiences.

Common sense suggests that this may be a "chicken and egg" question. At the secondary level, an entry attitude may be more relevant, to the extent that it affects course choices, energy invested in course work, participation in math clubs or competitions, etc. This suggests the outcome attitude--the result of prior math experiences--may be especially important for high math ability students in the later elementary/early middle school years, since these attitudes would form the basis for the entry attitudes which students would bring to later, more specialized math opportunities. If students have not found satisfaction, felt encouragement from significant others in the face of challenge, formed motivational study habits, or come to believe in the importance of mathematics during the later elementary to early middle school years, it may be less likely that they will select or engage in challenging courses, when the choice is available, even if they could be successful. Of the studies included in Table 3 (above), only three address any outcome affective variables, and each has serious limitations as noted (Ewers and Wood, 1993; Hughes, 1985; Swiatek and Lupkowski-Shoplik, 2000). Thus, the empirical base of evidence about this issue for this group is nearly nonexistent.

This study attempted to fill this gap in the empirical literature, and to do so with more detailed, empirically sound data than any extant studies offer. Because of the policy regarding acceleration in the district where this research took place,[4] the study examines attitude variables in the absence of this systemic response to mathematical ability. The results will not be able to show that acceleration leads to increases in positive attitudes, as has been suggested in the secondary level research. However, it is possible that the results may show that lack of acceleration in mathematics is associated with decreases in attitude toward that subject, and less favorable attitudes than those found in the general population. This, in itself, would be a significant finding, as it would bolster the case for providing acceleration to older elementary students by examining data which addressed an area of the affective domain rather than achievement issues which lie in the cognitive domain.

The instrument used to measure attitude toward mathematics is the same one used in an earlier study (described below), an adapted form of the Fennema-Sherman Math Attitude Scales (Fennema-Sherman, 1976) (adapted F-S). The original form of this instrument has been used in many empirical studies of math attitude (c.f., Hyde et al, 1990). As will be detailed in the following section, an earlier study conducted in the same school district as the current study established that, in that school district, middle school students have consistently high scores on each of the attitude sub-scales measured by the adapted F-S. Therefore, any attitudes for the high math ability students which differ from the rest of the cohort by a statistically significant degree would be of interest, and should be informative to the research on this population.

The study improves the field in several ways. In light of the paucity of empirical research on the attitudes of gifted elementary math students, a study using the Fennema Sherman Math Attitudes Scales, which examines nine different sub-constructs of attitude toward mathematics, substantially increases the knowledge about this topic in the research literature. It also aligns with existing empirical literature about both the general population elementary students, and the high math ability secondary students, filling the currently existing gap with a study of similar variables. Because the study was conducted across an entire school district, the "n" is large enough to enhance generalizability of the results. And finally, because of the design of the study, the issues of self-selection which may have seriously biased the few other studies reviewed above are not present.

<u>Study Demographics</u>

The subjects in the current study were drawn from a single school district, a middle to upper middle class suburb of a large Midwestern city. The total student population over the seven years inclusive of the study cohorts has increased incrementally, from 11,300 to 12,000 students; each grade level cohort is generally between 800 and 900 students. The district has thirteen elementary (K-5) buildings, divided fairly evenly between buildings of less than 300 students (2 teachers per grade), buildings of 300 and 420 students (2 or 3 teachers per grade), and buildings of over 450 students (4 to 5 teachers per grade). The district attempts to maintain class sizes in the elementary level below 25, though occasional classes as large as 28 may occur, especially in the upper grades. Four middle schools (grades 6-8) had between of 600 to 750 students each during the years of the study; the "typical" count for a given year and building was in the middle to high 600's. The middle school attendance boundaries are relatively mixed--i.e., eight of the thirteen elementary schools feed students to more than one middle school. Thus, between grades four and seven, students in the study would encounter significant group mixing due to both changes in year-to-year teacher assignments and the realignment of elementary building cohorts at matriculation to middle school.

The district's records on ethnic distribution of students during the years of the study show that 5 to 7% were African American students, 4 to 6% were Asian students, less than 1% combined were Hispanic, American Indian or Alaskan Native students, and the remaining 85 to 90% were Caucasian students. The traditional Midwestern mix of students has changed over the last decade, as increasing numbers of immigrant families have moved to the area, due to job transfers and as a result of political unrest, particularly in the Middle East. The proportion of bilingual students (with over 60 languages spoken in students' homes) steadily increased from about 4 % in 1990 to 18% in 2000. From 1994/95 to 1997/98, the years in which the current study cohorts were identified, the increase was from 5% to 14%. Students who do not yet have a working knowledge of English are exempted from the district testing programs, so they are not represented in the sample of the current study. However, some portion of the students in the current study would be considered as having working, though limited, English skills; this data is not reported in the district's test reports.

Socioeconomic status of students in the district is generally high. The percentage of free and reduced lunch applications for the years of the study was stable at approximately 5%, with actual participation (i.e., percent of eligible participants actually buying under the program) fluctuating between 15 and 20% (i.e., less than 1% of students enrolled in the district regularly participate in the free or reduced lunch program.) Several buildings contributed the vast majority of these: three elementaries and one middle school recorded applications of between 10 and 20% during the years in question. In these buildings, participation rates fluctuated between 20 and 45% (i.e., in those buildings, between 2 to 9% of enrolled students regularly participate in the free or reduced lunch program.)

The figures reported here are for the district as a whole; the information on ethnic, bilingual or socioeconomic status of the gifted sample was not available from district records, though students from all these classifications were represented in the sample.

Sample Selection

Math Attitude Study in Middle School

As part of a general school improvement effort, an adapted version of the Fennema Sherman Math Attitude Scales (Fennema-Sherman, 1976) (adapted F-S) was administered in one of the district's four middle schools for three subsequent academic years to all students in the school, beginning in 1994/95 (see Appendix). This yielded data from the classes of 1999, 2000, 2001, 2002, and 2003. (In the current study, all cohorts will be described by their projected year of high school graduation.) This part of the study will hereafter be referred to as the baseline study, unless specific cohort years are mentioned for clarity. Thus, a data set was created which gave longitudinal data for the classes of 2000, 2001 and 2002. Table 4 displays the administrative design of the baseline study.

Table 4

Matrix of Administration Design for Adapted Fennema-Sherman Scales

Cohort by Graduation Year	99	00	01	02	03	N
Year of Administration						
94/95	8[a]	7	6			603
95/96		8	7	6		581
96/97			8	7	6	630

[a] Indicates grade level of students at time of administration

By cross-referencing the data set (using student numbers) with third grade Comprehensive Test of Basic Skills (CTBS) scores available in district test record files, it was possible to disaggregate the group by selected percentile cut-scores on the CTBS. It was also possible to compute mean attitude scores for a specific grade, supported by up to three cohort years of data. This data set provided baseline data for the students in this school district, showing student attitudes toward mathematics from the sixth through the eighth grade. It was thus able to be used for comparison to the data which were subsequently gathered about high ability students from subsequent graduating classes from the time they were in the fourth grade, as described next.

Math Ability Study In Elementary School

Based on the Miller et al. (1995) report of the successful use of the SMPY model with elementary aged students, in 1997 and 1998 the district conducted a pilot administration of the School and College Ability Test-Intermediate level (SCAT-Q, refer to endnote 1) to the fourth graders who scored at least at the 95 %ile on their third grade California Test of Basic Skills-Total Math (CTBS-TM). The pilot administration of the test was silent: Because it was done to gather district level, not individual student data, individual student scores were not reported to either teachers or parents, nor used for any individual student educational planning. This decision was made in response to administrators' belief that teachers provided adequate challenge through the use of general enrichment strategies which could challenge even the most able students. There were also concerns that if parents believed their students obtained high scores, they would pressure both the students and the district for more intense math instruction.

The Intermediate level of the SCAT-Q is normed for the sixth through ninth grade. The Spring, seventh grade norms were used in accordance with the recommendations in Miller et al. (1995): Students scoring above the 65%ile on the out-of-level norms should be given access to acceleration in addition to significant amount of enriched math instruction. The breakdown of students found at each of these critical cut points in the pilot administration is shown in Table 5.

Table 5

Students Scoring Above Critical Cutpoints In Pilot SCAT-Q Project

Class	N of cohort[a]	CTBS-TM ≥ 95%ile[a]	Percent of cohort	SCAT-Q > 65%ile [b]	Percent of cohort
2005	835	194	23.2	41	4.9
2006	889	170	19.1	41	4.6
TOTAL	1724	364	21.1	82	4.8

[a] third grade testing

[b] fourth grade testing, using out-of-level norms for Spring, seventh grade[5]

Despite the strong evidence in the literature that acceleration is the most effective response to advanced math ability, the district's longstanding policy (refer to endnote 4) effectively prevented acceleration in mathematics before the seventh grade, except in the most unusual circumstances (e.g., a sixth grade student moving in from another district would not be required to repeat course work documented on their records). This policy had strong administrative support, based on a philosophical belief in equity and the desire to avoid labeling or sorting elementary aged students by ability (cf. Oakes, 1985; Slavin, 1987). Enrichment of the regular mathematics curriculum was believed sufficient to meet the needs of all high ability math students until they reached the end of sixth grade. At that time, advanced students (determined by a combination of teacher recommendations and teacher-made cumulative tests) were recommended for placement in a pre-algebra class as seventh graders. Most of these students then took algebra I as eighth graders.

Teachers up through the sixth grade were encouraged (though not monitored or required) to provide enrichment lessons for all students, including the practice of "compacting out." "Compacting out," if permitted by the teacher, occurs when students take the textbook publisher's end-of-chapter 20-item test before studying the chapter. Those scoring above 90 or 95% were "compacted out" to enrichment lessons for the duration of the chapter instruction. As noted above, research shows that the use of enrichment strategies is minimal in most elementary classrooms (Delcourt et al., 1994; Westberg et al., 1993). Classroom teachers deal with many competing pressures, and a wide variation in classroom climate, teaching skill and style would be expected when considering the dozens of staff across the district at the grade levels in question. Thus, the access of high ability students to even the minimal enrichment as defined in the research literature would have to be considered sporadic--i.e., irregular, infrequent, idiosyncratic--and not systematically documented or assessed.

In accordance with the district's policy, these students' (05/06 cohort) math instruction was carried out in mixed ability classes, supplemented with enrichment at teacher discretion, until the selection was made of those who would take pre-algebra as seventh graders. This situation provided an ideal set of circumstances for a naturalistic longitudinal study of attitudes towards mathematics in this group of highly able students, when that ability was not accommodated in a systematic way over several years which include transition to middle school.

Measurement Instruments

Fennema Sherman Mathematics Attitude Scales

The instrument used to assess mathematics attitudes in the baseline study is an adapted version of the Fennema-Sherman Mathematics Attitude Scales (hereafter, F-S) (1976). It was originally conceptualized as a "measure of important, domain specific attitudes which have been hypothesized to be related to the learning of mathematics by all students and/or cognitive performance of females" (p. 1). The manual reports that content validity was established through a series of item analysis and factor analytic studies conducted in a middle class suburban/rural Midwestern high school (N= 367). From a preliminary bank of 173 potential items, the final version of the survey consisted of 108 items, divided into nine sub scales. All subscales are scored so that higher scores indicate a more favorable attitude. The nine subscales are:

- Confidence--beliefs in one's ability to learn and perform well on mathematical tasks;
- Success--degree to which students attribute mathematics success to positive or negative consequences;

- Usefulness--student's beliefs about the usefulness of mathematics;
- Effectance Motivation--persistence in engaging in problem solving activities;
- Math Anxiety--apprehension toward mathematics;
- Male Domain--mathematics is perceived to be appropriate to males; and
- Mother; Father; Teacher (three scales)--student's perceptions of each of these significant other's attitude or confidence toward them as learners of mathematics.

The correlations among eight of the sub scales are reported, disaggregated by gender; differences between the genders were reported as non-significant. Of the 56 resulting r values, 39 are below .5. The two highest r values were .66 (between male's confidence in learning mathematics and effectance motivation in mathematics), and .65 (between females's view of their mother's and their father's beliefs of their ability in mathematics). The generally low correlations support the assertion of the authors that each sub scale measures a construct which is somewhat related, but largely independent of the others. The ninth sub scale, Anxiety is the single exception, with a correlation (in an earlier phase of the scales' development) of .89 with the Confidence scale. The authors elected to retain it as a separate scale, however, due to the perceived interest in the field for this particular construct. The split-half reliabilities were computed and reported in the manual; reliabilities for the nine scales ranged from .86 to .93.

Each sub scale could be used alone or in any combination. Each sub scale is assessed by twelve prompts: Six are phrased in a positive way, and six in a negative way. The student responds on a five-point Likert scale. The range of raw scores for each scale is 12 to 60; higher scores indicate more positive attitudes toward mathematics. Descriptive statistics, developed for groups of ninth through twelfth grade high school students in the Madison, Wisconsin area in the spring of 1975 (N=1428), are also reported in the manual for each of the nine sub scales. Since its development, the F-S has been utilized in numerous studies reported in the literature (c.f., Fennema and Hart, 1994; Hyde, Fennema, and Lammon, 1990).

The adaptation of the F-S was made by the district committee which conducted the original district survey in 1994/95. As explained by the district Coordinator of Testing and Measurement, one of the original district study authors, (personal communication, February, 1998), although the F-S was deemed to be the best available instrument to measure math attitudes, it was believed that a 108 item survey was too lengthy for middle school students. Additionally, it was believed by the committee that there was strong potential for negative public relations if students were required to respond to negatively phrased items like "I don't think I could do advanced mathematics," "My father hates to do math" or "My teachers think advanced math is a waste of time for me." The committee reviewed the item mean scores (reported in the manual, 1976), and because there was very little overall difference between the means of the negatively vs. positively phrased items, the decision was made to adapt the survey by using only the positively phrased items. Thus, the adapted Fennema Sherman which was used by the district in its three-year study is a 54 item survey, in which each of the nine scales is assessed by six positively phrased prompts.[6] The five Likert scale choices range from "Strongly disagree" (1 point) to "Strongly agree" (5 points). Each adapted scale, therefore, has a raw score range of 6 to 30, with higher scores indicating more positive attitudes. (For the Male Domain scale, the higher score indicates a less gender stereotypical view of mathematics. The authors note this is believed to affect the motivation of females to study mathematics.)

The adapted F-S was administered to the fourth and fifth graders in small groups, by the district's coordinators of gifted assigned to each of the thirteen elementary buildings. Directions were read to students, and included assurances that their answers were being used for research purposes only, and that there were no right or wrong answers. The fifty-four prompts were then read to the group, while students answered on a Scan-tron sheet. Questions about the survey process, unfamiliar vocabulary, etc., were allowed, though almost none were reported by the gifted coordinators. At the middle school level, the administration was conducted in the regular math classes by the math teachers. This was similar to the format used in the baseline study, and was preferred by the middle school administration as being less disruptive of routines or suggestive of "who is gifted." Most teachers preferred giving the written survey to students, though some chose to administer the questionnaire orally. Though many middle school students completed the survey, the Scan-tron sheets of the students who had participated in the SCAT-Q testing as elementary school students were identified, and were the only sheets actually scored.

The School and College Ability Test

The Quantitative section of the School and College Ability Test (SCAT-Q) (1996) (refer to endnote 1) was used in the Miller et al. (1995) Appalachian project, and was therefore chosen for the pilot testing of the high ability math students in the fourth grade. The Quantitative section of the test is 50 items long. Students are allowed 20 minutes to complete the test. Each item presents two values (e.g., A. the area of a rectangle with a length of 3, and B. the area of a rectangle with a length of 5.) The students do not need to compute an actual answer, but rather must be able to see the relationships accurately enough to determine one of four possible choices: Answer A is greater, answer B is greater, A is equal to B, or not enough information is given to decide. (This last option would be the correct answer for the example just given). The problems represent a mix of arithmetic (including all fractions), number theory, geometry, algebra, statistics and graphing, and common measurement. The Intermediate Level, Form X was used; the selected out-of-level norms were the Spring, seventh grade, in accordance with the recommendation in the Miller et al. study. All SCAT-Q percentile scores reported in the current study are based on these out-of-level norms.

The SCAT-Q was administered to students in small groups by the coordinators of the gifted assigned to each building. Students were told they were "testing the test"-- i.e., that no grades or other important decision was dependent on their individual score, but that the teachers were interested in finding out if the test gave useful information about good fourth grade math students. The students were accustomed to being involved in various enrichment activities with the gifted coordinator, and seemed to experience the SCAT-Q testing as another enrichment option.

Comprehensive Test of Basic Skills

The district used the complete battery of the Comprehensive Test of Basic Skills, Fourth edition (CTBS/4) (1990) to gather achievement data for children in the third grade. The Total Math (TM) score is a composite score developed from scores on the Mathematics Computations and Math Concepts and Applications subtests. The *Test Coordinators's Handbook* describes Mathematics Computations as "a test of traditional computation operations performed with whole numbers, decimals, fractions, integers, percents, and algebraic expressions, and including skills of applying the standard order of operations, mental math and estimation." The Mathematics Concepts and Applications subtest is described as "a test of problem solving encompassing the broad range of strategies and critical thinking skills students use in confronting problems, and includes traditional one- and two-step word problems, non routine problem solving items, estimation and pattern recognition skills, and concept formation skills" (p. 12). National percentile norms, comparing students to other third graders in the national norming sample, were one of the metrics chosen for reporting by the district. It was the one used in determining the students for the current study.

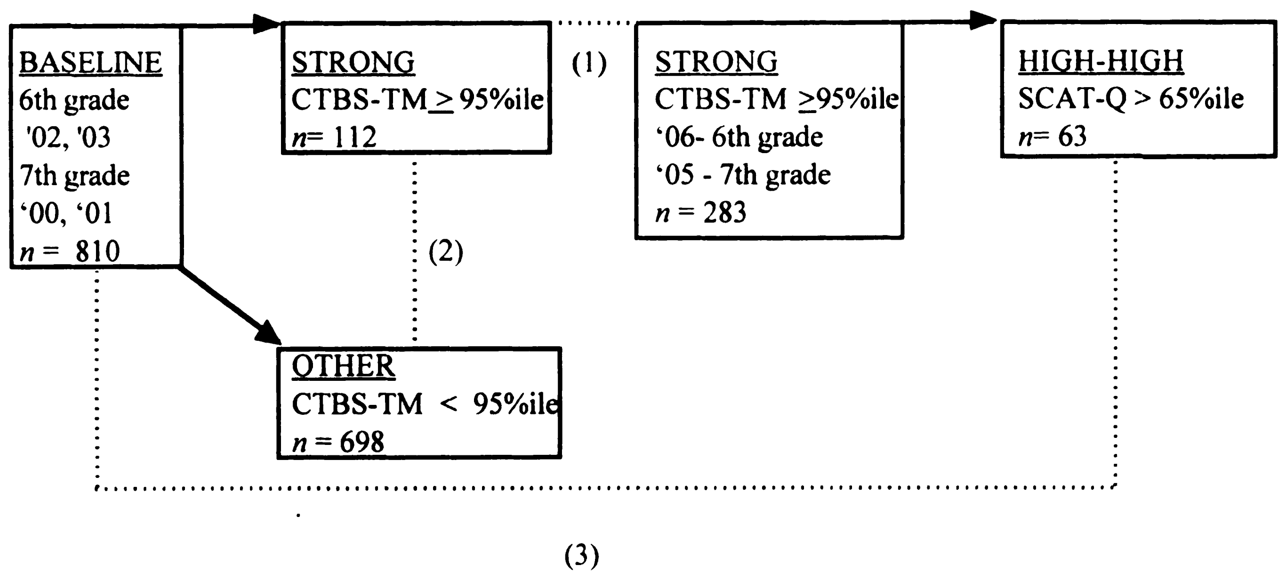
Research Design and Methodology

For clarity in the following narrative sections, the groups are designated as follows. Baseline students are those students who were part of the original three-year study conducted by the school district. All baseline students are sixth or seventh graders, from the entire range of ability. "Strong" students are those who scored at or above the 95%ile on the third grade CTBS-TM; all students in the 05/06 cohort are, *de facto*, "strong" students. Within the 05/06 cohort, "middle-high" are those who scored at or above the 95%ile on the third grade CTBS-TM but below the 65%ile on the SCAT-Q, while "high-high" are those who scored at or above the 95%ile on the third grade CTBS-TM and also scored below the 65%ile on the SCAT-Q.

The general research design was a longitudinal explanatory study, a special case of nonexperimental quantitative research (Johnson, 2001). Math ability (defined by cut scores on achievement measures) was the independent variable and math attitude (as measured by the adapted Fennema Sherman Math Attitude Scale) was the dependent variable. Multivariate analysis of variance (MANOVA) tests were used to determine the effect of math ability on math attitudes. First, the study explored this relationship for sixth and seventh grade students taught in regular classroom settings (see Figure 1). Additionally, changes in the math attitude of the middle-high and of the high-high groups were measured for each grade between the fourth and seventh grade (see Figure 2). This phase of the study used a panel design, with data for the same children collected at approximately one year intervals. To avoid violation of the assumption of independence

Figure 1

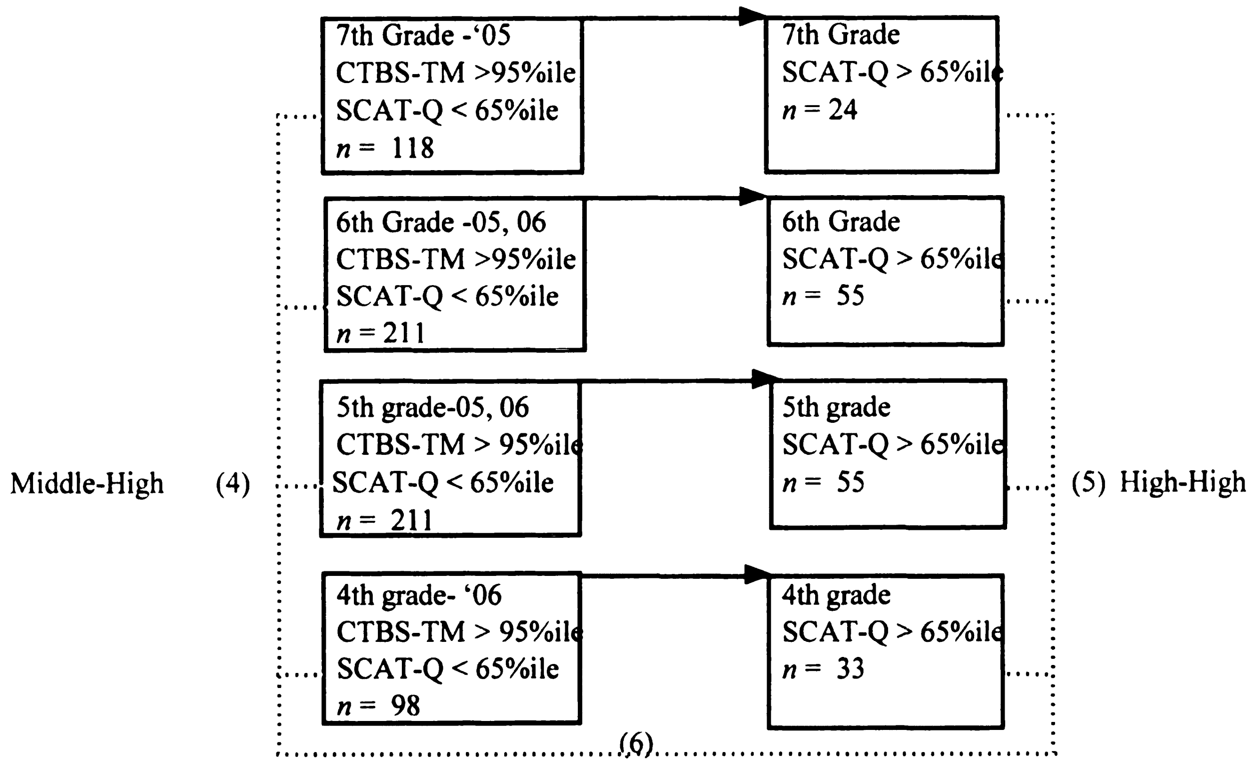
Comparisons at Sixth and Seventh Grade Between Baseline Groups and Class of 05/06



ns are numbers actually used in MANOVAs; attrition from original sample not included.
Numbers in parentheses refer to research questions.
Arrows indicate disaggregations of samples.

Figure 2

<u>Comparisons Between Sequential Grade Levels and Between Aggregate Groups for Classes of 05/06</u>



*n*s are actual numbers used in MANOVAs; attrition from original sample is not included

Numbers in parentheses refer to research questions. Arrows indicate disaggregations of samples.

of data, the test for this analysis was a doubly multivariate repeated measures MANOVA. Each child's performance was compared with his or her own score in the subsequent year, for each sub scale. Gender differences for selected research questions were also examined.

The alpha for both sets of research questions was set at $p \leq .10$ for the two-tailed MANOVAs, seeking a $p \leq .05$ in the direction deemed to show negative attitudes. The rationale for selecting this standard is as follows: The basic issue is seeking evidence to avoid a Type II error, resulting in no action taken when action (e.g., accelerating students) would be warranted. Accelerating students to ameliorate or prevent negative attitude would be warranted if student attitudes were shown to be harmed by the current (mixed ability) arrangement. If the attitudes of students in the highest ability aggregates were found to be either unchanged or improved in comparison to their peers, then reasons to change the current practice would be less compelling. By extension, this would suggest that there is little reason to be concerned when able math students are not accelerated in mathematics.

Question 1 also used an effect size criteria, $d = (m_H - m_L) / s$, as this was the same criteria used in the original district report on the baseline data. Cohen points out, "the only concern about magnitude in much psychological research is with regard to the statistical test result and its accompanying p value, not with regard to the psychological phenomenon under study " (1991, p. 155). Using p values as the only criteria for either the baseline study or the psychological variables being studied in question 1 (the similarity of the baseline and sample groups) would substantially increase the risk of committing a Type I error. Common sense suggests that Cohen's medium effect size ($d > .5$ for N of at least 85, in tests comparing independent means between groups of similar N), which represents "differences visible to the naked eye of a careful observer"

(p. 156), is a more salient statistic for this question. The differences between two cohorts from the same stable school setting, separated by less than five years, are not likely to be great, despite the findings of statistical significance. Documenting this through the use of the effect size statistic both followed the reporting in the baseline study, and made eminent sense when considering the question which was posed.

For the subsequent research questions, however, the more stringent MANOVA standards were used exclusively, rather than the effect size criteria. This was done specifically because differences on the attitude scales may exist but may NOT yet be "visible to the naked eye of a careful observer" (Cohen, 1991, p. 156) during the grade levels in question. The essential issue driving the current research was concern for the early development of attitudes toward math, so seeking evidence of such a difference even before it may become "visible" in ways which would impact on math-related decisions was of critical importance. (Effect size was also reported for questions 2 and 3, to give consistency for all the group comparisons represented in Figure 1. However, group size differences were substantial and, in one instance, did not meet Cohen's minimum suggested N [1991, p.157], so effect size was not considered the statistic of choice for these two questions.)

For questions 1, 2 and 3 (Figure 1), a cross-sectional design was planned, drawing the comparison groups of combined sixth and seventh graders from two sources: The baseline group, and the students involved in the pilot SCAT-Q testing which was done when students in the classes of 2005 and 2006 were in the fourth grade. For each of the two comparison groups, the combined sixth-seventh graders were drawn from cohort years in a manner such that no single child was included more than once. In the baseline group, this meant the sixth grade was drawn only from the 02 and 03 cohorts, and the seventh grade only from the 00 and 01 cohorts. For the 05/06 group, the sixth grade was drawn from the 06 cohort and the seventh grade from the 05 cohort.

The baseline data set was also disaggregated by ability levels, divided into students who scored at or above the 95%ile on their third grade CTBS-TM scores (strong students) and students who scored below the 95%ile on that test. In the 05/06 cohorts, students who scored at or above the 95%ile on the third grade CTBS-TM were targeted for inclusion in the pilot SCAT-Q testing. Thus, only students who were in the district as third graders and scored at or above the 95th %ile on the CTBS-TM were included in the 05/06 sample. This group was further disaggregated into the high-high (above the 65%ile on the SCAT-Q), and middle-high (below the 65%ile on SCAT-Q) students. The sixth and seventh grades were combined in all disaggregations for questions 1, 2 and 3.

This disaggregation was also used for research questions 4, 5 and 6 (Figure 2), though for these questions, the grade levels were separated. The adapted F-S was administered for only three years in sequence, so the '06 cohort contributes the only fourth grade data, and the '05 cohort contributes the only seventh grade data. Table 6 summarizes the *n* and source-years for each sub-set of data.

Table 6

Number of Students in Grade Level Groups
Disaggregated by Critical Cutpoints

Group	N	> 95%ile CTBS-TM [a]	> 65%ile SCAT-Q [b]
Baseline Grade Six/Seven	1190	148	
2005/06 Grade Four	1724	364	82
2005/06 Grade Six	1770	321 [c]	63 [c]

[a] Third grade scores

[b] Fourth grade scores

[c] Students who were in original pool of third graders, only.

Limitations

Sample Attrition and Curricular Stability

One of the limitations of longitudinal research is the problem of attrition. Some students leave to attend private or parochial school (there are several elite private schools in the area), while others leave due to family moves. The greatest attrition in the current study occurred among the high-high group, between the fourth and sixth grade, when approximately 25% of these students left the district. (This is higher than the district's normal attrition rate, which is approximately 25% between the fourth and seventh grade.) Because the study design used third grade CTBS-TM scores as the initial selection criteria, it was not possible to incorporate students of comparable ability who enter the district after grade 3. District test records, however, suggest that the math ability of each cohort is stable, and that the mix of mathematical abilities in classrooms is similar across grade levels and cohort years. To the extent that the range of abilities in classrooms may be one strong element in the "affective features of school and classroom," described above as Anderson and Bourke's (2000) second stage of affective variables, the classroom ability mix could be considered stable over the years being studied.

A potential limitation would be a change in district math curriculum and/or teaching methods. For the cohorts in question, the district's math curriculum was stable (texts, philosophy, etc.) for each of the years the students were studied. Individual teacher styles and classroom climate would be expected to vary, however, and the instructional staff during the years of the study experienced typical turnover due to reassignments and attrition. The district valued and supported general improvement of teacher skills through ongoing professional development, but no particular emphasis was placed on mathematics during the years of the current study. Thus, instructional style and classroom issues would be considered as random error factors in the current study.

Time Series Analysis

Because the current study design is longitudinal, and involves data comparisons at several different points over the course of the study, questions may arise related to time series analysis. Ostrom (1990) defines two general categories of time series regression analysis: the nonlagged case and the lagged case. In the first, both the dependent and independent variable are observed at the same point in time, while in the lagged model, the dependent variable is related to past values of either the dependent and/or independent variable. Both cases are based on a regression model, wherein one variable is hypothesized to have a relationship to the other variable which can be expressed in a mathematical formula of the general form, $Y= a + bX + e$. Ostrom defines the difficulty that occurs because the error term must be considered as correlated from one point in time to another--i.e., it is likely to be biased at each point in the data gathering by the outcome of earlier activities in the relationship being studied, resulting in potentially severe effects on estimated coefficients. This skews, in an ultimately predictable way, the relationship defined by the formula.

While the cut-scores being used to disaggregate the groups in the current study would be considered independent variables, the current study did not seek a mathematical relationship between them and the attitude variables, but rather proposed them as a way to operationalize the factor of high math ability. Because the current study neither proposed nor sought a regression formula, the issue of time series regression analysis is moot. Furthermore, because it is a naturalistic study, it is quite probable that any bias in the error factors will be similar from one point to the next, and attempting to eliminate these could, in the long run, destroy the very thing the study is seeking to establish--i.e., do the many real-world variables which impact on student attitudes toward mathematics play themselves out differently for students believed to have extremely high ability, than they do for students with more typical ability?

Scale Quality

Though the overall quality of the original Fennema-Sherman scale is strong, there are some limitations to be considered due to the adaptation in the current study. Anderson and Bourke (2000) summarize issues which impact on scale quality. They group these issues into four major categories: Communication Value, Objectivity, Reliability and Validity (p. 90). While acknowledging that few self-report instruments meet all of the requirements, Anderson and Bourke suggest that the more that are met, the greater the confidence in the instrument. The original Fennema-Sherman Math Attitude Scales (1975) met these quality issue requirements in nearly every way.

However, the adaptation of the scales created some potential difficulties. Discussion and explanations of how these were accommodated for the current research follows.

Communication Value. The F-S is an instrument which is nearly twenty-five years old at the time of the current study. While a few items contained colloquialisms which might be considered dated (e.g., "gotten shook up" to mean stressed), most items were nonetheless considered straightforward and culturally neutral. Students taking the scales, whether administered orally or in written form, appeared to have little difficulty understanding the purpose of the survey or the meaning of the prompts, and responding differentially to the items.

Objectivity. Objectivity addresses issues of coding rules and errors, and of scorer competence. The five-point Likert scale, ranging from "Strongly Agree" to "Strongly Disagree," was clearly presented in the instructions, and students did not express any confusion about using the Likert scale. Students were told that the middle choice, "Undecided," could represent either a neutral opinion, or an actual lack of decision, and that students should answer all prompts with their first impression, rather than deliberating over a response. Scan-tron sheets were checked, before scoring, to eliminate marking errors (stray marks, duplicate or omitted answers for a prompt, incomplete surveys, etc.) Stray marks were erased, and other coding errors were found on less than 1% of the surveys, well below the acceptable error limit of 10% suggested by Anderson and Bourke (2000). The surveys were electronically scored, effectively eliminating the issue of scorer competence.

Reliability. Anderson and Bourke (2000) describe four issues pertaining to the examination of affective scale reliability. Each of these four issues is addressed here:

a. Comparability of sample groups: The misalignment between the original F-S scale norm group (high school students) and the current target group (middle school and upper elementary students) renders the descriptive statistics reported in the manual useless for the current purpose. However, these data will not be used in compiling the statistics for the current study. For purposes of this research, the means and standard deviations developed using the district's original study will be considered as local norms.

b. Internal consistency of each scale: The F-S manual (1976) includes a report comparing the mean scores for each item. In most instances, each positive item had a corresponding negative item (e.g., "Math tests don't especially worry me" and "I really dread math tests.") A comparison of the means of the "pairs" of positive and negatively phrased items, conducted by the designers of the original (baseline) study (C. Mahalek, personal communication, April, 1998) revealed that the data on negatively phrased items were highly consistent with the data on the positively phrased items. The study designers decided, based on that perusal, that the reliability as determined by the split-half correlations reported in the manual (ranging from .86 to .93 for the nine scales) would not be seriously compromised by the use of only the positively phrased items.[6]

c. Degree of stability: The mean scores reported in the original district study (Appendix) showed only modest variability (as measured by effect size, *d*) across grade levels or cohort years, which suggested stability in the adapted scales when used for middle school students. As explained by the county's educational resource center Director of Research and Evaluation, one of the original committee members, because the report was intended for a lay audience who would be naive about statistical reporting, the committee decided to use an easily understood raw-score-points criteria to determine significance (personal communication, July 9, 2001). Each of the many cohorts, grades and gender aggregates revealed standard deviations on each scale between 3.9 and 5.3 points. A difference of two raw score points between group means, therefore, was a practical way to explain effect sizes of approximately .3 to .5, a criteria which suggested small to medium differences (Cohen, 1991). It was further believed by the original committee that more stringent statistical measures, which might have shown significance, were misleading--i.e., statistical significance between the groups in question did not necessarily reveal educationally important differences. The effect size criteria resulted in a report of great stability among the students over the three years of the baseline study, which matched practitioner observations of the groups taking the survey. This reasoning is congruent with Cohen's, which posits that a medium effect size of $d = .5$, "represent(ing) an effect likely to be visible to the naked eye of a careful observer " (p. 156), is a much more salient statistic than p for many psychological phenomenon (see discussion above, page 61).

d. Whether equivalence with a similar scale is relevant to the particular purpose of the scale: The broad use of the Fennema-Sherman scales in the research literature on attitudes toward math suggested that an adaptation would give an equivalent scale.

Validity. Adaptation of the scales had the potential of seriously impacting on validity in several ways. Even under ideal circumstances, "With respect to [establishing] validity...do what you can. If two or three pieces of information relevant to the validity of the scale are offered and if these pieces support the validity of the scale, we believe there is satisfactory evidence of validity" (Anderson and Bourke, 2000, p. 91) .

There are strong split-half reliability estimates (ranging from .86 to .93 for the nine sub scales) as reported in the F-S Manual (Fennema and Sherman, 1976). Though impacted by the shortening of the instrument, the strength of the original split-half reliability suggests that reliability estimates of the adapted version would still be high. For example, the Mother's influence scale had the lowest reliability estimate of the nine scales, at .86. Using the Spearman-Brown formula, the reliability estimate for the adapted version with half the number of items would be .75. Robinson, Shaver and Wrightman (1991, referenced in Anderson and Bourke, 2000, p.111) consider this to be evidence of extensive reliability. The validity coefficient in this instance would be approximately .87--a strong validity estimate.

Another piece of validity evidence would be the correlation between the scale scores and measures of other characteristics thought to be related to the attitudes in question. The baseline study reported statistical significance using a Pearson Test of Significance*(sic)* at the .05 level for a number of subscales and students' most recent math grades (e.g., high Confidence scores and students who earned A's). (See Appendix , p. 10). No actual measures were reported.

The choice to use only positively phrased items could impact on validity in two ways. Meuller (1985) suggests that, for attitude measures, "we can *approximate* the content validity model by including a wide variety of of positive and negative opinion statements about the attitudinal object...(though) there is no statistical index of content validity. The process must simply be documented" (p. 63) (italics in original). He further states, "...internal consistency...can also supply evidence of construct validity. It does not constitute a complete validation. ...(but) by the very nature of the attitude-scale construction process (Likert and Thurstone, at least) some amount of *content* validity is assured. Content validity and internal consistency are often used in combination in attitude scale validation. The combination of content validity plus internal consistency supplies at least minimally acceptable evidence of construct validity for attitude scales" (p. 71). Selecting only positively phrased items could impact on both the (presumed) content validity, as well as the internal consistency, and thus might also skew the construct validity. As noted above, however, the examination of the item means carried out by the baseline study committee revealed little variation between the pairs of negatively and positively phrased items, which suggested high internal consistency for the positive items alone (refer to endnote 6).

Anderson and Bourke (2000) suggest that the selection of only positive prompts may create a potential for greater acquiescence (tendency to answer positively), which would impact on validity. Once again, however, the high internal consistency as shown by the similarity of item response scores between negatively and positively phrased items reported in the F-S Manual (Fennema and Sherman, 1976) suggested this factor may not impact heavily on this particular survey. Anderson and Bourke suggest other strategies to further control for the tendency to acquiescence. These are shortening the instrument and maintaining random order of sub scale prompts. Both these suggestions were incorporated in the adaptation.

A final issue related to validity is the use of the scales with an age group different than the original norming sample, which could impact on the content validity. According to Meuller (1985), the opinion of expert judges is one method for establishing content validity. The items which were used in the adapted scale were reviewed by the baseline study committee which included the district Coordinator of Testing and Measurement, the Director of Research and Evaluation at the county resource center, and a local district middle school counselor. All believed the prompts were understandable and would yield variable responses from middle school students. The district Math Coordinator, the district Coordinator of Testing and Measurement, and the district Coordinator of Gifted and Talented Services reviewed the items for use with able fourth and fifth graders, and believed they would be understood and yield variable answers by this younger group.

In summary, then, the validity of the adapted scale appeared to be "extensive... based on multiple methods of empirical validation with statistical support (e.g., differences and/or correlations in right direction) [and] some attention paid to instrument development and judgmental validity" (Anderson and Bourke, 2000, p. 111).

Research Questions

Two sets of research questions were planned for the current study.

The first set of questions compared the baseline data to the 05/06 cohorts, to establish possible comparability between the two cohorts and to seek evidence of possible differences in attitude between the full cohort and either the strong math students or the high-high group. (Figure 1, page 58, illustrates the pattern of comparisons, with each research question indicated by the numbers in parentheses.) The strong students in the baseline group were found through a review of district test records. The research question for this comparison were:

1. For students scoring at or above the 95%ile on the third grade CTBS-TM, is there a mean score difference at the sixth-seventh grade level on any of the nine sub scales of the adapted F-S between the cohorts in the baseline study and the '05 /'06 cohort? (Two tests for significance will be applied: MANOVA with alpha set at $p < .10$, and effect size with a cut off set at $d >$ .5, as used in the original baseline study.)

If, as was predicted, the null hypothesis was sustained for this question, the assumption was made that the attitude data of the full cohort for both groups would be statistically comparable, meaning both groups are similar with regard to the issues addressed in the nine sub scales.

A second research question determined if there were significant mean score differences in attitude at the sixth and seventh grade level between the strong students and the rest of the cohort. This comparison could only be made with the baseline data. The research question was:

2. Is there a mean score difference in the baseline study at the sixth-seventh grade level on any of the nine sub scales of the adapted F-S between the students who scored below the 95%ile on the third grade CTBS-TM, and the students who scored at or above the 95%ile?

An additional research question examined if the sixth and seventh grade mean scores on the adapted F-S differ between the high-high students and the full cohort of the baseline study. The research question was:

3. At the sixth-seventh grade level, do mean scores on any of the nine sub scales of the adapted F-S differ between the baseline study students, and the group of students in the 05/06 cohorts who scored below the 65%ile on the SCAT-Q?

The second set of research questions determined if there were developmental patterns of change in attitudes about math between the fourth and the seventh grade, for students in the '05/'06 group (Figure 2, page 60 illustrates the pattern of comparisons, with each research question indicated by the numbers in parentheses.) The attitudes of the middle-high, and of the high-high groups were examined for each year, as well as the total change between fourth and seventh grade. The research questions were:

4. Is there a change in attitude toward math between the fourth grade and seventh grade, as shown by each year's mean scores on any of the sub scales of the adapted F-S, for the students whose third-grade CTBS-TM scores are above or equal to the 95%ile but below the 65%ile on the fourth grade SCAT-Q scores?

5. Is there a change in attitude between the fourth grade and the seventh grade scores, as shown by each year's mean scores on any of the sub scales of the adapted F-S, for the students whose fourth grade SCAT-Q scores are below 65%ile ?

Last, the possible difference between the middle-high and high-high groups at each year was examined, as a way to determine if there is a critical period during these years when the opinion of the high-high group may begin to diverge from the rest of the cohort. The research question for this will be:

6. Is there a difference in attitude, as shown by each year's mean scores on any of the sub scales of the adapted F-S, at either the fourth, fifth, sixth or seventh grade level between the students whose third-grade CTBS-TM scores are above or equal to the 95%ile but below the 65%ile on the fourth grade SCAT-Q, and those whose SCAT-Q scores are above the 65%ile?

Empirical results on these questions helped answer the question of whether failing to provide adequate challenge to elementary aged precocious mathematics students may contribute to a decreased level of enthusiasm and interest in mathematics. Empirical results also contributed to the sparse body of knowledge about the attitude variables for mathematically precocious children in elementary school.

Results

The current study involved a series of questions (see Figures 1 and 2, page 58 and 59). While each was relatively independent of the other, the overall pattern which was of eventual interest. A brief discussion and interpretation accompanies the reporting of the results for each question. The Conclusions chapter integrates the results and addresses the implications and generalizations for the several questions considered holistically. In the reporting of the results, as much as possible the term "cohort" is used to refer to students by their graduating year(s), while the term "group" designates an identified disaggregate of one or several graduating years. For each question, first the null hypothesis is stated, and the results of the MANOVA reported. When ANOVA results revealed significant results for individual subscales, these are reported next. Exploration of the gender results related to each question is the last data reported, followed by the brief discussion and interpretation relevant to that question's results.

Correlations

Two sets of correlation data were computed, preparatory to answering the research questions. The first set was an inter-scale correlation, to assure relative independence of each of the measures for the younger-than-norming-sample cohorts. The results were congruent with the inter-scale correlations reported by Fennema and Sherman (1976) for the original norming sample. (They reported results by gender, only, and found no differences between the genders for inter-scale correlations.) The correlations among the scales for both the original and the current sample are shown in Table 7. As with the original norming sample, the correlations for the current study cohorts are generally low.

Table 7

Correlations Among Fennema-Sherman Math Attitude Scales

For Original Norming Sample and Current Study Sample

Scale	Confidence			Mother			Father			Success			Teacher			Male Domain			Usefulness			Anxiety		
Group[a]	M	F	C	M	F	C	M	F	C	M	F	C	M	F	C	M	F	C	M	F	C	M	F	C
Confidence																								
Mother	.38	.48	.40																					
Father	.34	.38	.38	.57	.65	.64																		
Attitude/Success	.37	.23	.39	.42	.33	.40	.35	.33	.38															
Teacher	.61	.64	.41	.45	.52	.50	.37	.46	.42	.39	.32	.43												
Male Domain	.23	.22	.12	.23	.25	.31	.18	.23	.28	.30	.34	.20	.38	.25	.22									
Usefulness	.46	.44	.41	.53	.58	.52	.51	.58	.52	.40	.36	.49	.50	.44	.44	.27	.25	.24						
Anxiety [b]	na	na	.76	na	na	.34	na	na	.32	na	na	.37	na	na	.42	na	na	.10	na	na	.40			
Effectance Motivation	.66	.61	.59	.37	.46	.37	.33	.37	.35	.40	.34	.53	.53	.50	.51	.23	.17	.14	.50	.55	.51	na	na	.62

[a] Original data reported by gender, only: M= norming males (n = 644); F= norming females (n =589) ; C= current study, both genders (N =1175)

[b] Correlations between Anxiety and other scales were not reported for the original norming samples

In both the original sample and the current study, the correlation between the Mother and Father subscales and between the Confidence and Effectance Motivation subscales was moderately strong, in the .60 range. The correlation between the Confidence and Teacher subscales was similarly moderately strong in the original sample, but was much weaker (.41) in the current study. The current study found a high (.76) correlation between the Anxiety and Confidence subscales, which aligns with the original sample's results. (Correlations between Anxiety and other subscales were not reported for the original norming study. Fennema and Sherman noted that the Anxiety and Confidence constructs were actually not distinct, but they decided to retain the Anxiety subscale because of the keen interest around this construct in the field.) A moderately strong (.62) correlation was found in the current study between Anxiety and Effectance Motivation.

As was true for the original norming sample, this suggested that for these younger age students using an adapted form of the scales, the scales were somewhat interrelated though each scale measures a different construct (with the exception of the Anxiety subscale, as noted above.) The similarity of the correlation coefficients between the two sample groups also supported the validity of the adapted form of the scales for use with younger age groups.

The second correlation was between math achievement as measured by CTBS scores, and each of the attitude scales. Each of the correlations between the groups of strong students and each of the scales is significant ($p < .01$). The correlation coefficients were nenetheless small, as follows: Confidence, .33; Mother, .15; Father, .11; Success, .15; Teacher, .20; Male Domain, .08; Usefulness, .09; Anxiety, .28; and Effectance Motivation, .15. Thus, a predictable, but weak, relationship appears to exist between math achievement and attitudes toward math for strong math achievers.

Comparability at the Sixth-Seventh Grade: Figure 1 Questions

Question 1

The research question was: For students scoring above or equal to the 95%ile on the third grade CTBS-TM, is there a mean score difference at the sixth-seventh grade level on any of the nine subscales of the adapted F-S between the cohorts in the baseline study and the 05/06 cohort? The null hypothesis was: For students scoring above or equal to the 95%ile on the third grade CTBS-TM, there is no mean score difference at the sixth-seventh grade level on any of the nine subscales of the adapted F-S between the cohorts in the baseline study and the 05/06 cohort.

Results. As explained above, two statistical tests were conducted to determine the degree of similarity between the two cohorts: A MANOVA with $\alpha = .10$ in alignment with the other MANOVAs in the current study, and an effect size test, $d > .5$, reflecting the criteria used in the baseline study. The result is shown in Table 8.

Table 8

Two Comparisons of Attitude Scales Between Baseline Strong Group And 05/06 Strong Group in Sixth and Seventh Grade

Scale	Mean B'line	Mean 05/06	*S.D.* (within group)	*F* d.f.(1, 381)	*d*
Confidence	26.31	27.00	3.65	2.87*	.19
Mother	26.72	27.32	3.64	2.36	.17
Father	26.14	26.90	4.40	2.40	.17
Success	25.85	27.85	4.09	19.70****	.48
Teacher	21.37	24.47	4.59	39.59****	.68♦
Male Domain	26.64	28.19	3.96	12.60****	.39
Usefulness	25.09	26.29	4.39	5.99**	.27
Anxiety	22.80	24.05	4.88	5.25**	.35
Effect. Mot.	20.39	22.41	5.37	11.25***	.37

Baseline $n = 110$; 05/06 $n = 283$

* $p < .10$; ** $p < .05$; *** $p < .01$; **** $p < .001$; ♦ $d > .5$, medium effect size

These results allow two different interpretations. The MANOVA found significant differences between the two cohorts, Wilks' $\Lambda = .863$, $F(1, 381) = 6.76$, $p < .001$. Examination of the data in Table 8 show the cohorts appeared very different from each other on seven of the nine scales when considering the ANOVA results for each scale. However, using the effect size criteria, the cohorts were as similar to each other as the cohorts in the baseline study were to each other, with the exception of the Teacher scale (see Appendix). The null hypothesis would be rejected if the more stringent MANOVA *F* values were used. To better align with the criteria used in the original baseline study, however, for this question the effect size criteria was used. Since the effect size is less than . 5 for all but one scale, the null hypothesis was sustained, meaning the two cohorts are essentially the same when comparing the strong students in each. By extension, it will be assumed that the full 05/06 cohort would be comparable to the baseline cohorts. This will allow for statistical comparisons planned for later questions.

Gender comparisons. The frequency count of each gender among the strong students was 266 boys and 211 girls (strong groups from all cohorts combined). Possible gender differences in the combined strong group was tested with a one-way MANOVA. Significant differences were found between genders, Wilks' $\Lambda = .812$, $F(1, 475) = 11.986$, $p < .001$. ANOVAs conducted on each of the nine scales showed significant differences on three subscales: Boys were more confident ($p < .05$), have less anxiety ($p < .05$) and held much more stereotypical views of math (Male Domain, $p < .001$).

Discussion. The strong groups from each of the cohort years held similar attitudes toward math, with the exception of the perception of their teachers where the 05/06 cohort was noticeably more positive. Since the smallest effect size is for the Mother and Father scales (i.e., reflecting values in the home), it seems reasonable to speculate that the more positive scores of the 05/06 cohort are, at least in part, a reflection of the milieu at school--a difference which may be most strongly revealed in the significant effect size difference for the Teacher subscale. A possible explanation is that, because the 05/06 cohort reflects views of students from all four middle schools while the baseline cohorts are from a single building, one teacher with a strong personality could unduly influence the views of the baseline students.

Although the differences between the cohorts are comparable to the differences found between groups and cohorts in the baseline study, it is none-the-less noteworthy that the 05/06 cohort had higher scores on every measure. This is not easily explained with available data. The higher scores of the strong 05/06 cohort may influence the comparison made in question 3. The other questions, however, should not be affected, as they involve comparisons only between aggregate groups from within a single cohort.

Among the strong students in all cohorts, the girls at the sixth and seventh grade had lower confidence and higher anxiety than the boys. However, the girls held a less stereotypical view of math than the boys, a somewhat ironic finding considering the lower confidence scores and the non-significant differences on the other six scales.

Question 2

The research question was: Is there a mean score difference in the baseline study at the sixth-seventh grade level on any of the nine subscales of the adapted F-S between the students who scored less than the 95%ile on the third grade CTBS-TM, and the students who scored above or equal to the 95%ile? The null hypothesis was: There is no difference in the baseline study at the sixth-seventh grade level on any of the nine subscales of the adapted F-S between the students who scored < 95%ile on the third grade CTBS-TM, and the students who scored above or equal to the 95%ile.

Results. A one-way MANOVA was conducted to determine if a difference existed between the two groups. Significant differences were found between the groups, Wilks' $\Lambda = .939$, $F(1, 806) = 5.788$, $p < .001$. The null hypothesis was rejected.

ANOVAs conducted on the nine subscale variables revealed that only three were significantly different. Table 9 shows the means, standard deviations, F and d between the groups for each of the subscales. (Effect size, d, was computed by using the within group standard deviation for the combined strong and other group: $\text{mean}_S - \text{mean}_O / s_W$)

Table 9

Results of ANOVAs Comparing Strong Baseline Students To All Other Baseline Students

Subscale	Mean Strong	Mean Other	*SD* Strong	*SD* Other	*F* d.f.(1, 806)	*d*
Confidence	26.31	23.77	3.49	4.72	29.335****	.55
Mother	26.72	26.02	3.16	3.83	3.311*	.19
Father	26.14	25.70	3.99	4.36	.951	.10
Success	25.85	26.13	4.38	4.18	.396	-.07
Teacher	21.37	21.80	4.36	4.64	.805	-.09
Male Domain	26.64	27.22	4.30	3.92	2.032	-.14
Usefulness	25.09	25.21	4.84	4.21	.077	-.02
Anxiety	22.80	20.87	4.73	5.04	14.111****	.38
Effect. Mot.	20.39	20.18	5.35	5.19	.159	.03

Strong $n = 110$ Other $n = 698$
* $p < .10$; **** $p < .001$

The mean scores revealed that the strong group held a more favorable attitude toward math for the Confidence, Mother, Father, Anxiety and Effectance Motivation subscale. Though the other group held more favorable attitudes for the Success, Teacher, Male Domain and Usefulness subscales, these four subscale differences were negligible. Considering $d > .5$ to be a difference which would be apparent to the eye of a careful observer, the greater confidence shown by the strong students would be the only area in which differences might be noticed. However, the MANOVA results indicated the presence of more subtle significant differences between these two groups by the middle school years.

Gender. The frequency count for each gender in the baseline cohorts was 415 boys and 395 girls. Gender differences for the baseline cohorts were tested in two ways. A one-way MANOVA was conducted to determine if there were significant gender differences among the full baseline cohorts. Significant differences were found between the genders, Wilks' $\Lambda = .800$, $F(1, 806) = 22.117$, $p < .001$. ANOVAs conducted on each of the nine subscales found significant differences on three: Confidence and Anxiety (ps < .001) favoring boys, and Male Domain ($p < .001$) with boys being more stereotypical in their view. For the second test, a two-way MANOVA was performed to see if there was an interaction between gender and strong student status. The interaction was not significant ($p > .50$).

Discussion. For this cohort of students, there was an overall difference in the attitudes toward mathematics between strong achievers and other students. Influenced primarily by their greater confidence and lower anxiety, and to a small extent by the perception of their mother's confidence in their mathematical skills , the strong students were more positively disposed toward mathematics than their less mathematically able peers. For the full baseline cohorts, girls were less confident and had higher anxiety levels about math, but held less stereotypical views about math than the boys. These findings parallel those of the strong students reported above, though the differences are more pronounced for the full baseline students. Being classified as strong or not strong was not related to a student's gender.

Question 3

The research question was: At the sixth and seventh grade levels, is there a difference in the mean scores on any of the nine subscales of the adapted F-S between the baseline study students, and the group of students in the 05/06 cohort who scored above the 65%ile on the SCAT-Q? The null hypothesis was: At the sixth and seventh grade levels, there is no difference in the mean scores on any of the nine subscales of the adapted F-S between the baseline study students, and the group of students in the 05/06 cohort who scored above the 65%ile on the SCAT-Q.

Results. A one way MANOVA was conducted to determine if there was a difference between the two groups. Significant differences were found between the baseline students and the high-high group (from the 05/06 cohort), Wilks' $\Lambda = .933$, $F(1, 869) = 6.889$, $p < .001$. The null hypothesis was rejected.

ANOVAs conducted on the nine subscale variables revealed that eight of the nine subscales were significantly different. Table 10 shows the means, standard deviations *F*,, and *d* for each subscale. (Effect size, *d*, was computed using the within group standard deviation for the combined high-high and baseline groups: $mean_h - mean_b / s_w$.)

Table 10

Results of ANOVAs Comparing High-High Students
To Baseline Students For Adapted F-S Subscales

Subscale	Mean H-H	Mean B'line	*SD* H-H	*SD* B'line	*F* d.f. (1, 869)	*d*
Confidence	27.89	24.11	2.65	4.65	40.416****	.81
Mother	28.08	26.11	2.86	3.75	16.526****	.52
Father	27.03	25.76	4.33	4.31	5.048**	.29
Success	28.13	26.09	3.81	4.21	13.901****	.48
Teacher	25.32	21.74	3.63	4.60	36.002****	.76
Male Domain	27.70	27.14	4.38	3.98	1.140	.13
Usefulness	26.79	25.20	3.99	4.30	8.142***	.36
Anxiety	24.83	21.13	4.18	5.04	32.024****	.80
Effect. Mot.	22.98	20.21	5.08	5.21	16.655****	.53

H-H = High High, n = 63; B'line = Baseline, n = 808
* $p < .10$; ** $p < .05$; *** $p < .01$; **** $p < .001$

The mean scores show that the high-high group at the sixth and seventh grade had a more favorable attitude toward math on every measure than did the baseline cohort, though the difference was not significant on the Male Domain scale. The magnitude of the difference was also revealed in the effect sizes, which ranged from moderate to strong for all subscales except Male Domain.

Gender. In the high-high group, there were approximately twice as many boys as girls (44 vs. 19). A one-way MANOVA was conducted to determine if there were gender differences in the high-high group. Overall, the gender differences were not significant ($p > .1$). However, the ANOVAs conducted on each of the nine subscales did reveal significant differences on three: Father, $p < .05$; Male Domain, $p < .05$; and Anxiety, $p < .05$. The mean scores for gender in the high-high group showed: girls have a stronger perception of their father's view of them as mathematically able than do boys and the girls feel less anxiety about math than the boys. As with other groups, the boys hold a more stereotypical view of math.

Discussion. As a group, the high-high students had much more favorable views toward math than their less mathematically able peers. At the sixth and seventh grade level, these students were highly confident, seemed to perceive the study of mathematics as a source of satisfaction and importance, and saw themselves positively in the feedback they receive from their significant others. To some extent, the higher scores of the high-high group are an artifact of the higher scores found for the 05/06 cohort in question 1. However, an examination of the means for the 05/06 cohort (see Table 8, page 78), and of the means reported here for the high-high group showed the high-high group had the stronger score in every instance. Thus, while the generally higher scores of the 05/06 cohort did attenuate the results somewhat, it was still apparent that the high-high students held much more favorable attitudes toward math than their peers. The boys held more stereotypical views of math, views which clearly aligned with their age peers, regardless of ability. High-high girls perceived particular support from their fathers, and were the least anxious about mathematics of all their peers.

Differences between sixth and seventh grade

The grade-to-grade differences for the strong group were tested in the second set of research questions (see Figure 2, page 60), using the data gathered from the 05/06 cohort. The original baseline study presented this data for the baseline cohorts (see

Appendix), but due to the use of the effect-size statistic, no significant differences were found. To provide a context for the data which will be reported in the second set of research questions, the possible differences between the sixth and seventh grade scores of the full baseline cohorts were tested. The null hypothesis was: There is no difference in mean scores for any of the nine subscales of the adapted F-S between the sixth and seventh grade.

A one-way MANOVA was conducted to determine if a difference existed between sixth and seventh graders. Significant differences were found between the groups, Wilks' Λ = ..868, $F(1, 806) = 13.463$, $p < .001$. The null hypothesis was rejected.

ANOVAs which were conducted on each of the subscales revealed significant differences for all subscales. Table 11 gives the means, standard deviations, F, and d for each subscale. (Effect size, d, was computed using the within group standard deviation: mean $_6$ - mean$_7$ / s_W.)

Table 11

Results of ANOVAs Comparing Sixth to Seventh Graders
For Baseline Cohorts

Subscale	Mean Sixth	Mean Seventh	*SD* Sixth	*SD* Seventh	*F* d.f. (1, 806)	*d*
Confidence	24.66	23.50	4.43	4.83	12.778****	.25
Mother	26.46	25.72	3.64	3.85	7.927***	.20
Father	26.15	25.33	4.17	4.44	7.431***	.19
Success	26.84	25.25	3.75	4.53	29.762****	.38
Teacher	23.25	20.04	4.23	4.42	111.054****	.70
Male Domain	27.43	26.81	3.73	4.22	4.983**	.15
Usefulness	25.76	24.56	3.87	4.66	15.866****	.28
Anxiety	21.72	20.48	4.94	5.08	12.310****	.24
Effect. Mot.	21.24	19.04	5.16	5.03	37.436****	.42

Note: Sixth grade was from 02/03 cohort; seventh grade was from 00/01 cohort
Sixth grade $n = 428$ Seventh grade $n = 380$
** $p < .05$; *** $p < .01$; **** $p < .001$

Except for the Teacher scale, the differences were small enough that they would not be individually revealed by the modest effect sizes (i.e., the test used in the original baseline study; see Appendix). Effect size is a convenient gauge for describing effects which are likely to be readily observed. However, because the intent of the current research was to discover nascent effects which might *not* be noticeable until they developed more fully, the more stringent MANOVA test was the test of choice for this comparison. The MANOVA results suggested that taken collectively, student attitudes toward math were significantly more positive in the sixth grade than in the seventh when considering all students in the cohort. These results must be understood with some caution, however. In order to maintain individuality of subjects, the comparison is actually of two different cohorts; results could be in part a reflection of cohort, not grade level, differences.

Developmental Changes In The Middle-high and High-high Groups: Figure 2 Questions

The questions which were examined next are those identified in Figure 2 (page 59). They were seeking evidence of differences between the middle-high and high-high students, on a year by year basis. Because SCAT-Q data was not available for the baseline cohorts, these questions examined only the students in the 05/06 cohort.

Question 4

The research question was: Is there a change in attitude toward math between the fourth and seventh grade, as shown by each year's mean scores on any of the subscales of the adapted F-S, among the students whose third-grade CTBS-TM scores are above or equal to the 95%ile but below the 65%ile on the fourth grade SCAT-Q scores? The null hypothesis was: There is no change in attitude toward math between the fourth and seventh grade, as shown by each year's mean scores on any of the subscales of the adapted F-S, among the students whose third-grade CTBS-TM score are above or equal to the 95%ile but below the 65%ile on the fourth grade SCAT-Q scores.

Results. A doubly multivariate repeated measures MANOVA was performed, comparing each of the three pairs of years: fourth to fifth, fifth to sixth and sixth to seventh. Table 12 shows the overall differences which were found for the middle-high group at each grade level comparison.

Table 12

Changes By Grade Level In Overall Attitude Toward Math
For Middle-high Students

Grade Level	Wilks' Λ	F	d.f.
Fourth to Fifth (n=98)	.823	2.13 **	9, 89
Fifth to Sixth (n=211)	.810	5.24 ****	9, 202
Sixth to Seventh (n=118)	.863	1.92 *	9,109

* $p < .1$; ** $p < .05$; ***$p < .001$; **** $P < .001$

Significant changes in attitude toward math occurred at each grade level for the middle-high students: Fourth to fifth grade, fifth to sixth grade, and sixth and seventh grade. The null hypothesis is rejected. The direction of the changes can be seen through an examination of the mean scores on each of the subscales. A repeated measures ANOVA was performed to test the significance of change for each of the subscales, for each of the contiguous grade level pairs. Table 13 shows the results of these comparisons.

Table 13

Grade Level Comparison of Middle-High Group
On Nine Subscales of Adapted F-S For Three Contiguous Grade Pairs

	Grade 4 to 5 (98)		Grade 5 to 6 (211)		Grade 6 to 7 (118)	
	Means	*F* d.f. (1, 97)	Means	*F* d.f. (1,210)	Means	*F* d.f.(1, 117)
Subscale						
Confidence	**26.82-27.06**	**.79**	**26.69-27.02**	**1.65**	27.14-26.78	1.06
Mother	**26.24-27.26**	**8.56*** **	**26.98-27.27**	**1.70**	**27.11-27.14**	**.01**
Father	**25.32-26.27**	**5.14** **	**25.84-26.77**	**9.67*** **	**26.25-26.75**	**1.43**
Success	**28.69-29.05**	**1.89**	28.33-28.09	1.25	27.96-27.41	2.34
Teacher	**24.84-25.32**	**1.17**	24.68-24.49	1.45	24.36-24.25	.07
Male Dom.	**27.65-28.42**	**4.26** **	**27.84-28.17**	**1.43**	**27.63-28.32**	**4.15** **
Usefulness	**27.01-27.14**	**.16**	26.69-26.43	.86	26.01-25.69	.78
Anxiety	25.23-24.87	.67	24.38-24.12	.52	24.22-23.97	.38
Effect. Mot.	24.70-24.70	0	23.78-22.75	10.61***	22.31-21.55	3.27*

Note: *n*s for each grade level pair appear in parentheses, and are the students who appear in each of the two contiguous grades being compared.
Fourth grade is only the 06 cohort, seventh grade is only the 05 cohort.
Bold indicates results which show improved attitude
* $p < .1$; ** $p < .05$; *** $p < .01$; **** $p < .001$

Gender. The numbers of boys and girls in this sample was essentially equal, for each of the three comparisons. A repeated measures ANOVA was performed to determine if there were significant gender differences for any of the contiguous grade pairs. For the students in the middle-high group, no significant gender differences were found in their overall attitude toward math for the fifth to sixth or the sixth to seventh grade comparison. However, significance was found for the fourth to fifth comparison (*F*

= 8.18 (9, 88), $p < .01$). An examination of the means showed that despite improvement in attitudes for the girls in every area from fourth to fifth grade, the middle-high boys generally held a more favorable attitude toward math in the fourth and fifth grade. Table 14 illustrates this data.

Table 14

Gender Based Means And Discrepancies For Middle-High Group
At Fourth And Fifth Grades

	Boys (48)			Girls (50)			Raw Score Discrepancy	
Subscale	4th	5th	Disc.	4th	5th	Disc.	4th	5th
Confidence	27.81	27.41	.40	**25.86**	**26.72**	**.86**	1.94 b	.69 b
Mother	**27.10**	**27.81**	**.71**	**25.42**	**26.72**	**1.30**	1.68 b	1.09 b
Father	**26.29**	**26.83**	**.54**	**24.40**	**25.72**	**1.32**	1.89 b	1.11 b
Success	**29.45**	**29.52**	**.07**	**27.96**	**28.60**	**.64**	1.45 b	.92 b
Teacher	**25.38**	**25.96**	**.58**	**24.34**	**24.70**	**.44**	1.04 b	1.26 b
Male Dom.	**26.50**	**27.20**	**.70**	**28.76**	**29.60**	**.84**	2.26 g	2.40 g
Usefulness	27.56	27.40	.16	**26.48**	**26.90**	**.48**	1.08 b	.50 b
Anxiety	26.75	25.85	1.10	**23.78**	**23.92**	**.14**	2.97 b	2.03 b
Effect. Mot.	26.06	25.18	.88	**23.40**	**24.24**	**.84**	2.66 b	.94 b

Note: *n*s are in parentheses .
g = discrepancy favors girls b = discrepancy favors boys
Bold type indicates change in a positive direction

Discussion. Students of very strong mathematical ability but not at the top of the cohort (called "middle-high" in this research), showed a generally positive view toward mathematics, and a view which, for the most part, increased in favorability through the fifth grade. A mix of positive and negative changes marked the significant shift ($p < .001$) which occurred when these students move from fifth to sixth grade. There were more negative than positive changes at the move to seventh grade, though all changes were small; only two reached statistical significance. The decline in the Effectance Motivation subscale reached a level of significance ($p < .01$) at the sixth grade, and again at the seventh ($p < .1$), suggesting that these bright students became less interested in pursuing problem solving activities when they move into middle school. Cumulatively, the decline seemed likely to be substantive by the time these students finished seventh grade.

There were some consistent increases in attitude for this group of students during these years, as well. The belief that math is a gender neutral subject increased every year, reaching a level of significance between both the fourth and fifth grade, and the sixth and seventh grade (ps < .05). There was also a continuous increase in the child's perception of both their mother's and their father's attitude toward them as math students, with the largest changes ($p < .01$) in the Mother subscale at the fourth-fifth grade, and the Father subscale at the fifth-sixth grade. A possible reason for the one year discrepancy may be the change from non-graded to graded report cards, between the fifth and sixth grade--perhaps good students believed their fathers were more interested in and proud of the "real" grades they received for the first time in the sixth grade.

Boys showed a mixture of increases and decreases in attitude toward math between the fourth and fifth grade, while girls attitudes toward math became more positive in every area. Boys, however, still held more favorable attitudes in every area than the girls, except one: Male Domain was the only area between the fourth and fifth grade in which mathematically able girls held a more favorable (i.e., find math gender neutral) view than mathematically able boys. (Though the overall gender differences did not reach a level of significance for the fifth to sixth and sixth to seventh comparisons, a similar pattern of girls having more gender neutral attitudes than boys existed during those years as well, despite girls having less favorable attitudes in other areas.)

Question 5

The research question was: Is there a change in attitude toward math between the fourth and seventh grade, as shown by each year's mean scores on any of the subscales of the adapted F-S, among the students whose fourth grade SCAT-Q scores are above the 65%ile? The null hypothesis was: There is no change in attitude toward math between the fourth and seventh grade, as shown by each year's mean scores on any of the subscales of the adapted F-S, among the students whose fourth grade SCAT-Q scores are below the 65%ile.

Results. A doubly multivariate repeated measures MANOVA was performed, comparing each of the three pairs of years: fourth to fifth, fifth to sixth and sixth to seventh. Table 15 shows the overall differences which were found for the high-high group at each grade level comparison.

Table 15

Changes By Grade Level In Overall Attitude Toward Math
For High-High Students

Grade Level	Wilks' Λ	*F*	d.f.
Fourth to Fifth (n = 33)	.554	2.15*	9, 24
Fifth to Sixth (n=55)	.791	1.35	9, 46
Sixth to Seventh (n= 24)	.574	1.24	9, 15

* $p < .1$

Significant changes in attitude towards math among the high-high group occurred only in the fourth to fifth grade transition . The null hypothesis was rejected for the fourth to fifth grade transition. The null hypothesis was retained, however, for the fifth to sixth grade transition, and the sixth to seventh grade transition. The direction of the changes can be seen through an examination of the mean scores on each of the subscales. A repeated measures ANOVA was conducted to test the significance of change for each of the subscales, for each of the contiguous grade level pairs. Table 16 shows the results of these comparisons.

Table 16

Grade Level Comparison of High-High Group
On Nine subscales of Adapted F-S For Three Contiguous Grade Pairs

subscale	Grade 4 to 5 (33) Means	F d.f. (1, 32)	Grade 5 to 6 (55) Means	F d.f. (1, 54)	Grade 6 to 7 (24) Means	F d.f. (1, 23)
Confidence	**27.64-28.55**	**4.53****	28.62-28.26	.67	28.33-28.04	.45
Mother	26.63-26.57	.01	**27.23-28.07**	**5.31****	**28.58-29.00**	**.80**
Father	**24.57-25.48**	**1.19**	**26.43-27.10**	**1.10**	**28.13-28.50**	**.51**
Success	**28.24-28.36**	**.08**	28.52-28.00	.95	**27.04-27.62**	**.43**
Teacher	25.06-24.09	1.90	**24.91-25.51**	**1.58**	**25.83-26.91**	**3.56***
Male Dom.	26.87-26.61	.08	**26.60-27.42**	**2.42**	**26.70-27.83**	**5.70****
Usefulness	26.78-26.12	.79	26.96-26.96	0	**27.33-27.81**	**.57**
Anxiety	26.63-26.33	.24	26.34-25.64	1.97	25.00-24.54	.43
Effect. Mot.	25.76-24.18	4.30**	24.96-23.67	3.57*	23.04-22.70	.1

Note: *n*s for each grade level pair appear in parentheses, and are the students who appear in each of the two contiguous grades being compared.
Fourth grade is only the 06 cohort, seventh grade is only the 05 cohort.
Bold indicates results which show improved attitude
* $p < .1$; ** $p <,05$; *** $p < .01$; ****$p < .001$

Gender. There were more boys than girls in the high-high group, by a ratio of slightly over two to one overall (each grade had a slightly different ratio). A repeated measures ANOVA was conducted to determine if there were significant gender differences for any of the contiguous grade pairs. No significant overall gender differences were found at any grade level pair between the high-high boys and high-high girls ($ps > .2$).

Though there was no overall difference, the means on the Male Domain subscale were examined to see if the same pattern was found; each year the girls had markedly higher scores than the boys for the Male Domain, meaning boys consistently held a more stereotypical view. At the fifth grade level, the girls mean score was 30, the highest possible score, while the boys at the fifth grade scored 24.95--a substantially lower score. Scores in all other areas were lower for girls than boys in the fourth, fifth and sixth grades. By seventh grade, however, the pattern changed somewhat, as the high-high girls began to surpass the boys in several of the subscales (though not, overall, reaching a level of significance). Table 17 illustrates this data for the seventh grade high-high group.

Table 17

Gender Based Means And Discrepancies For High-High Group At The Seventh Grade

Subscale	Boys (17)	Girls (7)	Raw Score Discrepancy
Confidence	27.82	28.57	.75 g
Mother	29.06	28.85	.21 b
Father	27.94	29.85	1.91 g
Success	28.17	26.28	1.89 b
Teacher	27.00	26.71	.29 b
Male Dom.	27.17	29.42	2.25 g
Usefulness	27.88	27.26	.62 b
Anxiety	23.94	26.00	2.06 g
Effect. Mot.	22.18	24.00	2.82 g

g = discrepancy favors girls   b = discrepancy favors boys
*n*s are in parentheses

Discussion. The most striking general finding for the high-high group was the high degree of stability of the attitudes. With some exceptions, these students evidenced little year-to-year change in their math attitudes from the fourth grade to the seventh, though cumulatively, there were some patterns which may be important.

While the students' Confidence increased to a statistically significant degree from the fourth to the fifth grade ($p < .05$), their Effectance Motivation showed a statistically significant decrease in the same year. The Effectance Motivation subscale shows a decline every year; reaching significance in the fourth to fifth grade and fifth to sixth grade comparisons ($p < .05$ and $p < .1$, respectively). A speculative, though plausible interpretation of this is that the high-high students feel extremely confident about performing the (for them) rather simplistic mathematical tasks required in the regular elementary school curriculum, but are wont to undertake more challenging problem solving tasks. Perhaps this is because they prefer "easy" to "harder" work, and the status which comes from being "the best or quickest math student" over the challenge of complex problem solving, or perhaps their experience with complex problem solving is so limited that they express lower interest out of naivete. Overall, the cumulative effect of the annual decreases in Effectance Motivation seems likely to be substantial for the high-high group by the time they finish seventh grade.

These students continue to perceive their parents as having favorable attitudes toward them as math students, and perceive their teachers similarly once they enter middle school. The Father and Mother subscales show small increases each year, with the Mother subscale reaching significance in the fifth to sixth comparison ($p < .05$). The Teacher subscale shows significant increase ($p < .1$) at the sixth to seventh grade comparison. The cumulative effect of the consistent changes on each of these three scales suggests a substantive net change for the high-high students between the fourth and seventh grade: These students consistently perceive the significant others as viewing them as highly capable math students. After almost no change from the fourth to the fifth grade, an increase occurred in the Male Domain area in each of the other two pairs of years, reaching statistical significance ($p < .05$) in the sixth to seventh comparison.

The gender discrepancy which has been apparent in previous studies of highly able math students is found here, as well. The ratio of two boys to every girl is not so great as in the studies of older students which used a more rigorous criteria to define high ability, but is nearly identical to that found in the Robinson et al. (1997) report on pre-school/primary mathematically gifted[5]. With regard to their overall attitudes, the high-

high boys and girls are quite similar during these years, except for the girls' persistently stronger opinion of math as a gender neutral subject. Though the overall difference in gender for the sixth to seventh grade comparison did not reach the level of significance, the subscale findings are nonetheless intriguing. A plausible explanation is that following several years of strong belief in the gender neutrality of mathematics, high-high girls, whose math ability would enable them to be successfully competitive with even the brightest boys, may "find their mathematical voice" resulting in decreased math anxiety and increased interest in pursuing complex problem solving around the seventh grade. Still, they are less certain than the boys that success in math is a socially desirable goal.

The results of the comparison of gender means for the high-high group must be considered tenuous for two reasons, however. First, the low n's, especially for the girls, make generalizability problematic. Secondly, the data are only descriptive, with no indication of statistical significance.

Question 6

The research question was: Is there a difference in attitude, as shown by each year's mean scores on any of the subscales of the adapted F-S, at either the fourth, fifth, sixth or seventh grade level between the students whose third-grade CTBS-TM scores are above or equal to the 95%ile but below the 65%ile on the fourth grade SCAT-Q, and those whose SCAT-Q scores are above the 65%ile? The null hypothesis was: There is no difference in attitude, as shown by each year's mean scores on any of the subscales of the adapted F-S, at either the fourth, fifth, sixth or seventh grade level between the students whose third-grade CTBS-TM scores are above or equal to the 95%ile but below the 65%ile on the fourth grade SCAT-Q, and those whose SCAT-Q scores are above the 65%ile?

A one-way MANOVA was conducted to determine if a different degree of change existed between the two groups, middle-high and high-high, at each grade level pair which was examined in questions four and five. The MANOVA was conducted by comparing the difference of the means of each contiguous grade level pair, between the two groups. (Tables 13 and 16 show the means of each contiguous grade level pair; the interested reader can readily derive the differences of the means.) Table 18 shows the results of these comparisons.

Table 18

MANOVAs Comparing Mean Score Differences Between Middle-high and High-high Groups For Contiguous Grade Level Pairs

Grade Levels	Wilk's Λ	*F*	*d.f.*
Grades 4 to 5	.8987	1.51	9, 121
Grades 5 to 6	.9711	.85	9, 256
Grades 6 to 7	.9599	.61	9, 132

Note: All *F* values are insignificant.

No comparisons reached a level of significance; the null hypothesis was retained.

Discussion The underlying issue which this question sought was if there were important differences in the developmental pattern of any changes of attitude, between the middle-high and high-high students. In general, the results did not show any such differences. With two exceptions, the comparison of relative changes on each subscale showed very little overall difference between the groups on a year-by-year comparison--i.e., the pattern of changes was generally parallel for the two groups. Examining the means for each group at each grade level (see Tables 13 and 16, pages 90 and 94) showed both groups have attitudes which were quite favorable on each subscale, with a general pattern of the high-high group holding a somewhat more positive attitude on most subscales.

The two exceptions were on the Male Domain and the Teacher subscale. Though both groups showed incremental increases between fourth and seventh grade on the Male Domain subscale (i.e., declines in stereotyping,) the high-high group held a slightly more stereotypic view of mathematics than their somewhat less able peers. This was true for all three contiguous years' comparisons.

The pattern for the teacher subscale was slightly more complex, as each group changed in an opposite direction for each comparison. The net result was that the high-high group showed a net gain (from the fourth to seventh grade) of 1.85 raw score points, while the middle-high group showed a net loss of .49. Though the differences between each level were statistically insignificant, the cumulative effect was substantive, especially since the two groups are not parallel (see Tables 13 and 16, pages 90 and 94).

Conclusions

Overview

The main hypothesis in this research was that students who have the highest abilities in math may experience a serious decline in their attitudes toward math when their abilities are neither challenged nor acknowledged through acceleration of their math curriculum during the later elementary and early middle school years. (The implications of this for instruction are discussed more fully below.) Each of the six questions in the research design attempted to provide empirical answers to different aspects of this issue. Table 19 gives an overview of results from questions two through five. (Question one established comparability between the baseline and 05/06 cohorts. Question six revealed no overall significance, so for purposes of clarity is not included in Table 19.)

The original hypothesis was not supported. The overall attitude toward mathematics for students with high math ability was robust, and appeared to withstand even several years of low-challenge instruction (see Tables 13 and 16, pages 90 and 94). In general, young, gifted math students had more favorable attitudes toward mathematics than their less able peers. Indeed, it appeared that the degree of math ability and the strength of the attitude have a direct relationship--the stronger the math ability, the more favorable the attitude (see Tables 9 and 10, pages 81 and 84), even after several years of less-than-challenging curriculum. Thus, concerns expressed in the literature (c.f., Ablard and Tissot, 1998; Waxman et al., 1996), that young high-ability students may become discouraged or disinterested in mathematics if not appropriately challenged, may be somewhat ameliorated based on the current research findings. However, there are important subtleties subsumed within this larger pattern, and these are examined next.

Though Fennema and Sherman (1976) report that each subscale is generally independent of the other, to facilitate the discussion of the conclusions, the subscales

Table 19

Overview Of Significant Results Of Comparisons For Questions Two Through Five

	Table 9	Table 10	Table 11	Table 13[a]			Table 16[a]		
	Strong to Other[b]	H-high to B'line[c]	B'line 6 to 7[d]	Mid-high Grade Pairs[d]			High-high Grade Pairs[d]		
SubScale				4-5	5-6	6-7	4-5	5-6	6-7
Confidence	**** S	**** H	**** 6				** 5		
Mother	* S	**** H	*** 6	*** 5	6	7		** 6	7
Father		** H	*** 6	** 5	*** 6	7	5	6	7
Success		**** H	**** 6						
Teacher		**** H	**** 6					6	* 7
Male Domain			** 6	** 5	6	** 7		6	** 7
Usefulness		*** H	**** 6						
Anxiety	**** S	**** H	**** 6	4	5	6	4	5	6
Effect. Mot.		**** H	**** 6	null	*** 5	* 6	** 4	* 5	6

a When significance is found, adjacent years are shown if direction of change was the same. If all three grade pairs showed the same direction of change, then all three pairs are reported even if significance was not found for any single pair.

b. S = mean score favored Strong group.

c. H = mean score favored High-high group.

d. Grade level which had the more favorable attitude is indicated

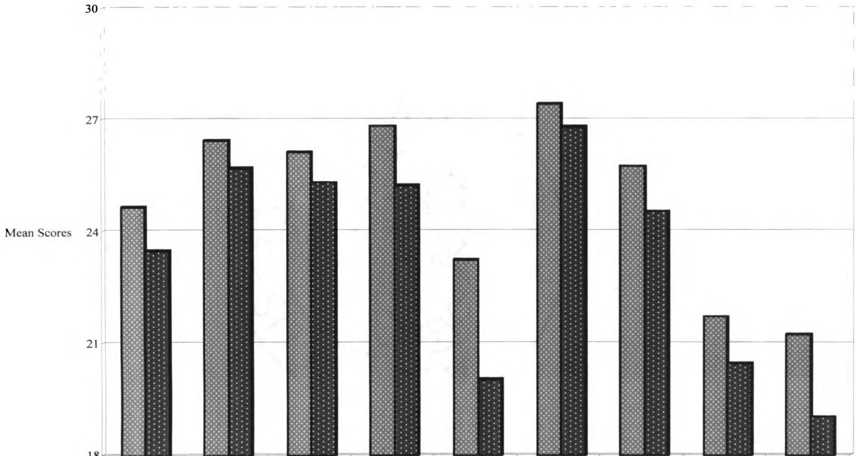
* $p < .1$ ** $p < .05$ *** $p < .01$ **** $p < .001$

were grouped into three clusters. The selection of areas and the subscales included in them was not mutually exclusive (two subscales appear in two areas), was admittedly arbitrary and even debatable. Based on the apparent real-world relationships of the qualities measured in each subscale, the three clusters were: Students' self-confidence as mediated by significant others; classroom milieu factors; and intrinsic factors.

Students' self confidence as mediated by significant others

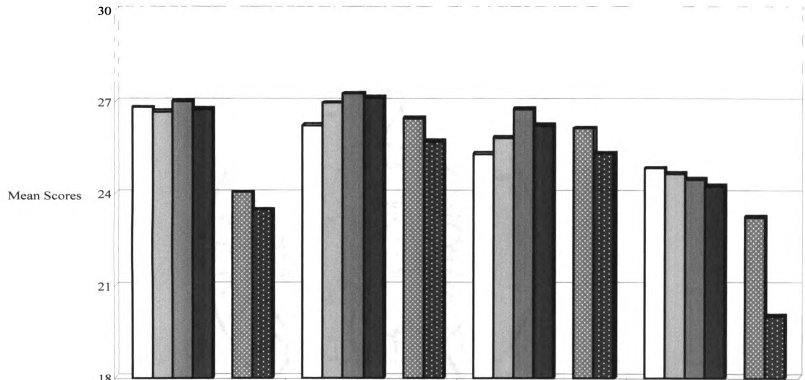
Previous research found students generally experience a small decline in attitude towards mathematics as they move through elementary school, a decline which accelerates dramatically as they move into junior high. This research also generally found the decline in attitude corresponded with an increasingly realistic self-perception of math competence (Eccles et al., 1993; Frome and Eccles, 1998; Wigfield and Eccles, 1992, 1994; Wigfield et al., 1991, 1997). The middle school decline for students of all ability levels was also found in the current study. Figure 3 illustrates this decline for all subscales (see also Table 11, page 87).

In marked contrast to their less able peers, however, there was no significant decline in students' confidence (the construct most closely aligned to the earlier studies' self-perception of competence) for either the middle high or high high groups (Figures 4 and 5; also see Tables 13, 16 and 19, pages 90, 94, and 102). This confidence was mediated by significant others, a point established in the studies by Frome and Eccles (1998) and Wigfield et al, (1997). The path models proposed by Eccles (1985) and Fennema and Peterson (1985) to explain achievement-related decisions also include the psycho-social environment (including, but not limited to, significant others) as a salient factor. Frome and Eccles, and Wigfield et al. surveyed parents directly and reported that mother's and father's perceptions significantly and increasingly corresponded with their children's self



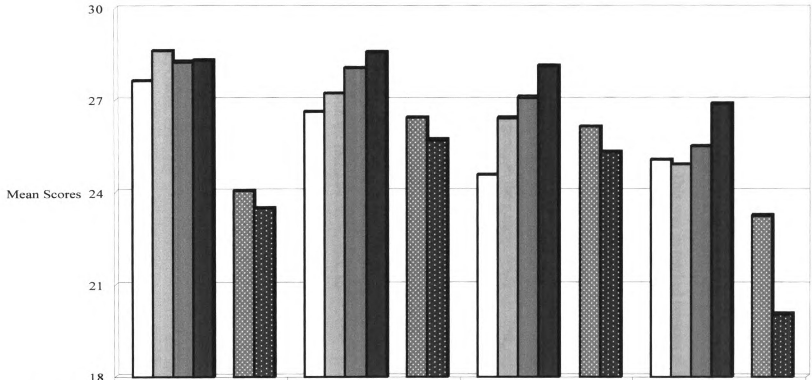
	Confidence	Mother	Father	Success	Teacher	Male Domain	Usefulness	Anxiety	Effect. Mot.
Sixth Grade	24.66	26.46	26.15	26.84	23.25	27.43	25.76	21.72	21.24
Seventh Grade	23.5	25.72	25.33	25.25	20.04	26.81	24.56	20.48	19.04

Figure 3. Comparison of baseline sixth and seventh graders for all subscales. Grid is drawn from Table 11, (p. 87) where standard deviations, significance values, and *n*s may be obtained.



	Confidence	Mother	Father	Teacher
Fourth Mid-high	26.82	26.24	25.32	24.84
Fifth Mid-high	26.69	26.98	25.84	24.68
Sixth Mid-high	27.02	27.27	26.77	24.49
Seventh Mid-high	26.78	27.14	26.25	24.25
Sixth Baseline	24.06	26.46	26.15	23.25
Seventh Baseline	23.5	25.72	25.33	20.04

<u>Figure 4.</u>Comparison of midde-high and baseline groups for significant others cluster. Grid data is drawn from Tables 11 and 13 (pp. 87 and 90) where standard deviations, significance values and *ns* may be obtained. Fourth grade is 06 cohort only; fifth and sixth grade are 05 and 06 cohorts; seventh grade is 05 cohort.



	Confidence	Mother	Father	Teacher
Fourth High-high	27.64	26.63	24.57	25.06
Fifth High-high	28.62	27.23	26.43	24.91
Sixth High-high	28.26	28.07	27.1	25.51
Seventh High-high	28.33	28.58	28.13	26.91
Sixth Baseline	24.06	26.46	26.15	23.25
Seventh Baseline	23.5	25.72	25.33	20.04

Figure 5. Comparison of high-high and baseline groups for significant others cluster. Grid data is drawn from Tables 11 and 16, where significance values, standard deviations and *n*s may be obtained. Fourth grade is 06 cohort only; fifth and sixth grade are 05 and 06 cohorts; seventh grade is 05 cohort only,

perceptions of competence as they moved through elementary grades and into middle school. Similarly, the current study found a pattern of increasingly positive perception by strong math students of both their mother's and their father's view of them as math students, with mothers holding a slightly more positive view than fathers for both subgroups of strong students. This was in marked contrast to the decline in the perception of either parent's view, held by the baseline students. Aligned with the earlier research, this finding would be expected for the very capable students: Parents held quite accurate views of their child's high math abilities, and appeared to communicate those views to their children during the years examined in this study.

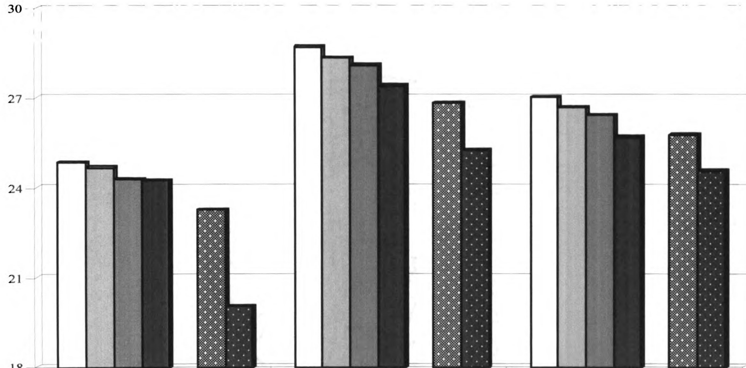
There was a divergence in the pattern, however, with regard to teacher views of students' math abilities between the earlier research and the current study. The findings by Wigfield et al. (1997) of teachers' views of student abilities are similar to their findings of parents' views--teacher views increasingly correspond to students' own competency beliefs and, along with parental views, mediate students' self perceptions of competency. As with the research on parents, the previous research surveyed teachers directly while the current research surveyed students' perceptions of teachers' views. In contrast to the findings regarding their parents, the current results showed that strong math students were receiving few messages from their teachers which would confirm their own competence/confidence. Perhaps teachers, influenced by egalitarian views espoused by many (c.f., Oakes, 1985) and Slavin, 1987), believed that acknowledging students who were at the top of the class would lead to decreased confidence or self-esteem for the students who were less able, and would establish a competitive, rather than cooperative, standard in the classroom. Perhaps the milieu of the middle school, with only 45 minutes a day spent with the math teacher, did not allow for enough connection with the teacher to facilitate teacher-student communication. Perhaps teachers chose to avoid communicating positively to students with advanced skills, in order to "keep them in their place" or to avoid dealing with parents who might bring pressure for more advanced work, if their students' abilities were acknowledged.

A decline in the Teacher scale was true for each year, for the middle-high students (see Figure 4 and Table 13, page 90). For the high-high students, a shift began at the sixth grade, and became a significant finding in a positive direction ($p < .1$) at the seventh grade (see Figure 5 and Table 16, page 94). Though not documented in the current study, the most likely explanation of this shift is that nearly all the high-high students (though not necessarily all the middle-highs) were almost certain to be identified through the district's screening method for placement into the seventh-grade pre-algebra class. They were also nearly all likely to be invited, early in their seventh grade year, to participate in the Midwest Talent Search , the SMPY-model testing program housed at Northwestern University. Thus, it would not be until the end of the sixth grade that these very capable students might first perceive (through placement in the pre-algebra class) definitive feedback from their teachers about their abilities as math students.

Classroom milieu factors

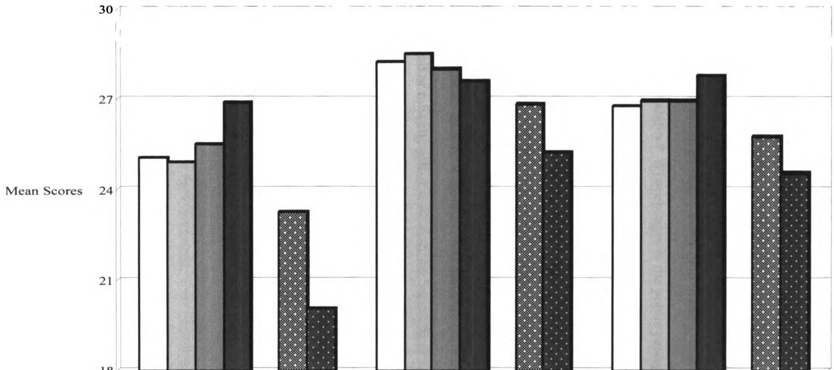
In isolation, this finding might hold little importance, given the strength of the findings regarding parents' views and the modest decline in student perceptions of teachers' views. But classrooms are multi-faceted, complex psycho-social environments in which teacher attitudes as well as peer values interact, and in which judgments are made based on a variety of factors (Smith et al., 1998). The social climate of the classroom milieu would constitute a significant portion of the psycho-social factors identified by Eccles (1985) and Fennema and Peterson (1985) in their path models of achievement decisions. Behaviors like enthusiastically answering a teacher-posed question, raising a probing or complex question about a math concept, or willingly helping a less able student could be considered everyday-achievement-oriented decisions. Behaviors like these would make math abilities socially visible in the classroom. A classroom milieu cluster of subscales could be composed of the Teacher, Success, Usefulness and Male Domain subscales. Everyday-achievement-oriented decisions would contribute to a child's perception of favor by the teacher, relative social standing, the usefulness of math in their lives, and the establishment of their public sex-role identity. Conversely, deciding to not act in a socially visible way would feel safer to students who may perceive the opportunity cost as too great on one or more of these factors, and thus avoid everyday-achievement-oriented decisions. Figures 6 and 7 illustrate the findings for these four subscales.

The belief that success in math is socially desirable declined for all groups during these years (see also Tables 13 and 16, pages 90 and 94). It seems logical that this might be due, at least in part, to a social climate influenced by the apparently bland and incrementally more negative message the brightest students (as well as other students) are receiving from their math teachers about their ability to do well in math. If teachers, who are the *de facto* arbiters of the institutional values about success in math, do not or cannot acknowledge in a public way that using one's abilities to the fullest is a desirable goal, then it is little surprise that students do not find this to be socially desirable. The pattern of the Usefulness subscale is similar--the lower the child's ability, the less the child perceives math as useful in their lives. It is especially intriguing to note that, in contrast to their peers, the high-high group's perception of usefulness spikes up at the seventh grade--mirroring the pattern of the perception of their teacher's view of them as math students.



	Teachr	Success	Usefulness
Fourth Mid-high	24.84	28.69	27.01
Fifth Mid-high	24.68	28.33	26.69
Sixth Mid-high	24.29	28.09	26.43
Seventh Mid-high	24.25	27.41	25.69
Sixth Baseline	23.26	26.84	25.76
Seventh Baseline	20.04	25.25	24.56

Figure 6. Comparison of middle-high and baseline groups on classroom milieu cluster. Grid data is drawn from Tables 11 and 13 (pp. 87 and 90), where standard deviations, significance values and *n*'s may be obtained.



	Teacher	Success	Usefulness
Fourth High-high	25.06	28.24	26.78
Fifth High-high	24.91	28.52	26.96
Sixth High-high	25.51	28	26.96
Seventh High-high	26.91	27.62	27.81
Sixth Baseline	23.26	26.84	25.76
Seventh Baseline	20.04	25.25	24.56

Figure 7. Comparison of high-high and baseline groups on classroom milieu cluster. Grid data is drawn from Tables 11 and 16, (pp. 87 and 94), where standard deviations, significance values and *n*s may be obtained.

The pattern of results for both the Success and Usefulness subscales suggested an important question: Might all students benefit, both in their belief of the social desirability of math success and their view of its usefulness for their lives, if schools were more public in their feedback to strong math students about their math competence? And might this change result in increased motivation to learn more difficult mathematics, regardless of one's ability? What changes in the public perception of math might occur if top math students were honored similarly to top athletes? The current research cannot answer these questions. Their logic is congruent, however, with Loveless' (1999a, 1999b) postulation that students lose when schools are organized to purposely hide evidence of high achievement, and that it is minority and disadvantaged students who lose the most--perhaps due to the relative lack of other feedback mechanisms in their psycho-social environments outside of school. Further research with a population of disadvantaged or minority gifted math students is needed to explore this dynamic more fully.

The final factor in the classroom milieu cluster is the view of math as a male domain. All three groups--baseline, middle-high, and high-high--showed scores which were statistically similar (see Table 19, page 102). However, Figures 6 and 7 revealed important contrasts between the strong students and the baseline group. The most striking was the direction of change--while the baseline group became more stereotypical in their gender views ($p < .05$), both the middle high and the high-high students developed significantly (ps $< .05$) less stereotypical beliefs from one year to the next. Indeed, the middle-high students held the least stereotypical views, though the difference does not reach significance. This finding seems to contrast with previous research, in which the brighter the student, the more stereotypical the view (Fennema and Hart, 1994; Hyde et al, 1990; Reis at al, 1996).

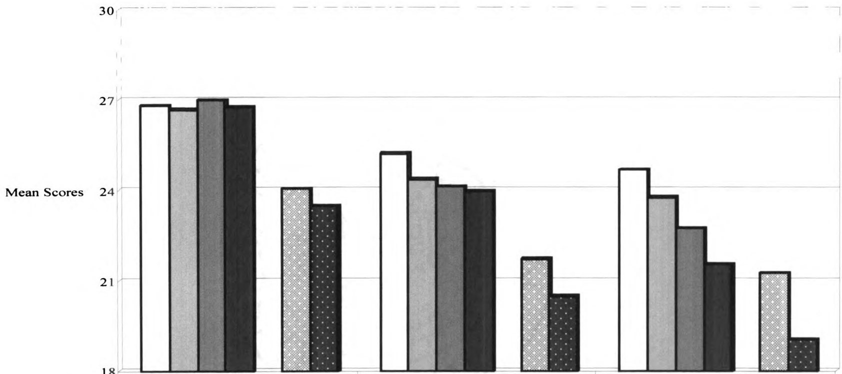
Though speculative, it is possible that the relatively high proportion of families with dual professional parent careers relying on math (e.g., accountants, physicians, engineers, and technicians) in this relatively affluent district could be an important factor contributing to the discrepant finding. In such families, math would be highly valued by mothers as well as fathers. Children from these families, both encouraged and predisposed to do very well in math, would likely constitute a disproportionate percentage of the strong math students. This possibility is congruent with the increases found for the strong students on the Mother and Father subscales, and also aligns with the finding from the literature that mothers are the main mediators of children's views of math (Jacobs and Eccles, 1992). It is not possible to tell from the current research if the parents' views of their children are because children of either gender are very competent at math, or if the children develop more gender neutral views because of their parents modeling and influence. In all likelihood, it is an interactive pattern.

An additional consideration may also help explain the discrepancy. The focus of the earlier studies (Fennema and Hart, 1994; Hyde et al, 1990; Reis at al, 1996) was on the differences between genders rather than ability levels, and may be somewhat misleading to the focus of the current research. (Those studies found high ability boys held strongly stereotypical views, which was also found in the current study.) Conclusions regarding the results of gender differences for the current study are more fully developed below.

Intrinsic factors

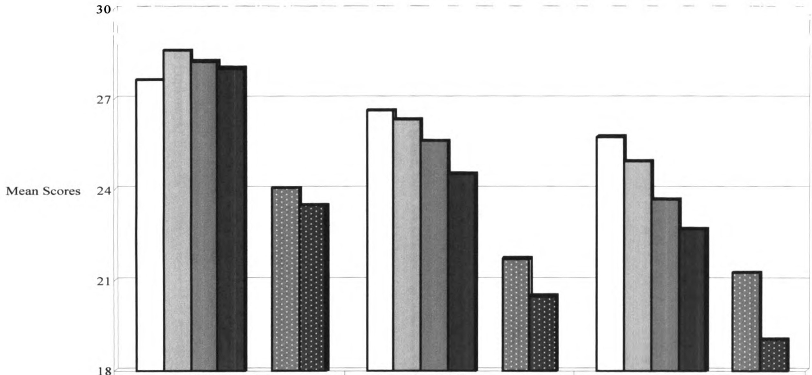
Because Anxiety was so strongly correlated with Confidence (r =.76 in the current study; r exceeding .80 in the original norming study [Fennema and Sherman, 1976]), the Confidence subscale will be considered with Anxiety and Effectance Motivation in a cluster dubbed "intrinsic factors." Each of the three factors in this cluster may be considered as originating within the individual's unique personality, rather than originating from extrinsic factors like significant relationships or the social milieu. Particularly in the case of confidence, it seems clear (as discussed above) that external factors impact on it. However, the impact appears to be iterative: Students may feel increased confidence following a pattern of success, and parents give encouraging and rewarding messages about that success, which leads to increased confidence and further success. As noted above, Frome and Eccles (1998), Jacobs and Eccles (1992), and Wigfield et al, (1997) found strong evidence of this mediational effect for students of all ability levels. The high correlations of Anxiety with Confidence suggests a similar pattern with an opposite vector may impact the Anxiety subscale. (The Anxiety scale is inversely related to felt anxiety--i.e., the higher the score, the less anxiety felt by the student.)

Examination of Figures 8 and 9 revealed an interesting pattern. While students of high ability were significantly more confident ($p < .001$) about their ability to learn math than their less able peers by the time they reached the upper elementary and middle school grades, it is a curiosity and perhaps cause for some concern that, for the most part, their confidence did not increase during these years to any significant degree. Neither the middle-high nor the high-high students showed any significant change (except for the high-high fourth to fifth shift--a shift that reversed the following year) (see Tables 16 and 19, pages 94 and 102). It seems reasonable to wonder if confidence for the two groups of strong students may fail to increase due to little opportunity for these very able students to exercise their ability--i.e., while they feel positively about consistently doing well in math class, they likely encounter few experiences in the regular curriculum which stretch their thinking and give them a chance to succeed at something which initially seemed hard to do, thus building confidence. Adding credence to this proposition are the small, persistent increases in anxiety (i.e., declining Anxiety scores). Though not reaching



	Confidence	Anxiety	Effec. Mot.
Fourth Mid-high	26.82	25.23	24.7
Fifth Mid-high	26.69	24.38	23.78
Sixth Mid-high	27.02	24.12	22.75
Seventh Mid-high	26.78	23.97	21.55
Sixth Baseline	24.06	21.72	21.24
Seventh Baseline	23.5	20.48	19.04

Figure 8. Comparison of middle-high and baseline groups on intrinsic cluster. Chart data is drawn from Tables 11 and 13, (pp. 87 and 90), where standard deviations, significance vaues and *n*s may be obtained. Fourth grade is 06 cohort only; fifth and sixth grade are 05 and 06 cohorts; seventh grade is 05 cohort only. Anxiety scores are inversely related to felt anxiety.



	Confidence	Anxiety	Effec. Mot.
Fourth High-high	27.64	26.63	25.76
Fifth High-high	28.62	26.34	24.96
Sixth High-high	28.26	25.64	23.67
Seventh High-high	28.04	24.54	22.7
Sixth Baseline	24.06	21.72	21.24
Seventh Baseline	23.5	20.48	19.04

Figure 9. Comparison of high-high and baseline groups on intrinsic cluster. Grid data is drawn from Tables 11 and 16 (pp. 87 and 94), where standard deviations, significance values and *n*s may be obtained. . Fourth grade is 06 cohort only; fifth and sixth grade are 05 and 06 cohorts; seventh grade is 05 cohort only. Anxiety scores are inversely related to felt anxiety.

significance in any single year, when considered in combination with the failure to gain confidence during these years, this may be suggestive of the development of a *faux* confidence. *Faux* confidence would occur when a student never encountered a challenge to their skill or understanding, and thus developed a view that math class was a place to show what they already knew, rather than a place to learn what they didn't know. This pattern seems especially probable in light of the relatively greater increase in anxiety felt by the high-high group. As math becomes more complex, the experience of "not already knowing" what is being taught may be creating greater anxiety for these students, while their slightly less able peers in the middle-high group may be more likely at an earlier age to have encountered--and "lived to tell the tale"--situations where they needed to expend some degree of effort to learn new material.

These concerns, while plausible, must be considered contextually: Clearly, in math up through the seventh grade, both the middle high and the high-high groups had significantly more favorable attitudes on these two intrinsic factors, as well as on the Effectance Motivation factor (all *p*s < .001, see Table 19) than their peers in the baseline group.

Even strong math students, however, experienced an incremental decrease in Effectance Motivation, a decline which reaches levels of significance at some grade levels between fourth and seventh grade (see Table 19, page 102). This parallels the findings by Swiatek and Lupkowski-Shoplik (2000), who reported an incremental decline between third and sixth grade in liking of math. They do not specify if students in their sample were involved in any accelerated math curriculum options in their home schools, though as first-time SMPY-model test takers screened for either verbal or math ability, it is likely many were not. This decline might be explained as a developmental phenomena--i.e., even very bright students are not immune to a decline in enthusiasm for academe.

However, the finding of any decline at all (albeit, less severe than for the less able students) is in contrast to the anecdotal evidence of improved attitude and increased interest and motivation resulting from accelerative options for this age level reported by Daniels and Cox (1988); Miller et al., (1995). The samples in each of these studies were self-selected, which would be likely to increase bias on this factor, and it is also possible that the subjective view of the researchers skewed their reported outcomes. If empirical data had been gathered on the attitudes of the students in these studies, perhaps the results would have paralleled those of the current study. However, this seems less likely than the following alternative explanation.

By default, the current study lends empirical support to the value of accelerative options. All three intrinsic factors could be affected. Rather than declining, Effectance Motivation scores may well increase with acceleration, as observed and reported by Daniels and Cox (1988) and Miller et al. (1995). It is also very possible that a concomitant increase in confidence and lowered anxiety would accrue as more difficult work was encountered and mastered, if acceleration were offered to students as early as their math abilities were apparent. The concerns expressed in the literature (c.f., Ablard and Tissot, 1998; Waxman et al., 1996) do appear to have some substance: Bright students who have not been allowed to accelerate do lose enthusiasm, do experience increased anxiety, and do not increase in confidence. Fortunately, even after the declines in these critical intrinsic factors, these students are still more favorably disposed toward math than their less able peers. Though the evidence is compelling, whether the declines can be directly attributed to lack of access to accelerative options is still a somewhat open question. Empirical research with young students who have been accelerated in response to their documented achievement is needed to bring the question to a close.

Gender issues.

The findings on gender differences in the current study were generally similar to those reported in the research literature. The frequency counts of boys and girls at each ability level revealed an essentially equal gender distribution for the baseline and the middle high group. However, in the high-high category, more that twice as many boys as girls were found. This is not as great a discrepancy as has been reported in the literature for the older students (Benbow and Arjmand, 1990; Benbow et al., 1996; Fox, 1976). Those studies, however, chose a more selective criteria: 97%ile on the out-of-level test vs. 65%ile on the out-of-level test used in the current study. Those researchers also reported a declining gender ratio as the ability level declined, which would be congruent with the current findings. The 2:1 ratio of boys to girls in the current study is strikingly similar to that reported by Robinson et al. (1996, 1997), in their studies of the pre-school/primary math gifted students. Thus, it appears that the gender achievement/ability gap continues unabated.

The most persistent finding on attitudes, for both the baseline group and the two groups comprising the strong students, was the consistently more stereotypical view of math held by the boys. This replicates the findings in the literature (Cramer, 1989; Eccles et al., 1993; Fennema and Hart, 1994; Hyde et al., 1990; Reis et al., 1996; Wigfield et al., 1997). However, some important differences on other dimensions between the current study and the reports in the literature were found .

Among the middle-high students, the overall gender differences at the fourth-fifth grade transition reached significance ($p < .01$) (see Question 4, page 88), and an even greater overall gender difference was reported for the strong students at the sixth and seventh grade ($p < .001$) (see Question 1, page 77), both favoring boys. The previous studies of attitudes of high ability students are limited, but differences in attitude of this magnitude have not been reported for elementary students. Indeed, those studies found little differences between the genders for any ability level (Eccles et al, 1993; Fennema and Hart, 1994; Hughes, 1982; Swiatek and Lupkowski-Shoplik, 2000). Though Wigfield et al., (1991, 1997), in their studies on the general population, reported stronger self-esteem for boys during these years in several areas, they noted that the least amount of difference was found in math, and that the reported differences may reflect response bias rather than genuine differences since boys tend to be more self-congratulatory than girls.

Even more intriguing was the increase in every area for girls at the fourth-fifth grade --but, an increase which none-the-less still found the girls substantially lower than the boys in every area except the Male Domain (see Table 14, page 91), and which was not maintained in subsequent years. A similar pattern of girls having generally less favorable scores than boys also existed for the high-high group (again, except for the Male Domain), though the gender differences for this group did not reach a level of significance at any grade level. However, the high-high girls at the sixth-seventh transition began to show stronger scores than the boys in several areas (see Table 17, page 96), with some differences being substantial. [7]

It seems that girls have learned to "talk the talk" of gender equity in math, but not yet "walk the walk," at least until the seventh grade for very bright girls. It is not easy to explain these findings which contrast with the more gender-neutral results of earlier research (Eccles et al, 1993; Fennema and Hart, 1994; Hughes, 1984; Swiatek and Lupkowski-Shoplik, 2000; Wigfield et al., 1991, 1997). While the contrast to earlier studies is clear, the paucity of earlier research on attitudes of highly able students makes generalizations problematic. Additional research is clearly needed to determine if, for example, there may be a critical developmental period for bright girls occurring around the fifth grade, or if the favorable changes for the high-high girls at the seventh grade can be taught or encouraged among girls of lesser ability. Would sixth grade be a prime year to implement girls-only math classes, to potentially allow for at least two years of growth in girls' attitudes toward math? Would a focused effort to bring fathers into the feedback loop (e.g., a daughters 'n dads math night, or smiley-face notes specifically sent to fathers for their daughters improved math grades) for girls of all abilities increase the likelihood that more girls would decrease their anxiety and increase or maintain confidence and enthusiasm for math through the pre- and early adolescent years? An answer to the question of how to create greater enthusiasm on the part of girls for mathematics is elusive, and it is disheartening to find results which seem to suggest that, though the girls held non stereotypical views, their other attitudes are less favorable than has been true in earlier studies.

Limitations and Further Research

Selecting the sample of this study from a single school district limits the generalizability to dissimilar populations. Though the district had a full range of students from all socioeconomic groups and included a broad cultural mix, the "typical" student in the study was middle-to-upper middle class, and came from homes reflecting traditional Midwestern values. Generalizing to other populations with consequential differences should be made with caution.

A related limiting factor was the institutional values of the school district, which are potentially biasing. Strong emphasis on high professional standards and ongoing staff development for teaching staff, and the clearly articulated expectation that all staff is responsible for teaching students of all ability levels creates a climate in most classrooms of acceptance for all. The district also provided (though did not require) substantial support for math enrichment, and many staff readily use these strategies on a regular basis. While not unique, these institutional values are often difficult to replicate in school systems having more limited resources or a higher proportion of "hard-to-teach" students. There may be a necessary and sufficient level of math enrichment which the students in this district were experiencing, which attenuated the possible negatives that failure to accelerate might cause. Thus, the relatively strong results for the students of strong math ability found in the current study may not occur in other settings. Replication research in other settings is needed.

A different sort of limitation is created by the instrument chosen for measuring attitude. While the Fennema-Sherman (1976) scale is broadly used in the research literature, and was the instrument used by the district previous to the initiation of the current study, it may fail to measure attitudes that could be important for the questions relative to the students of high math ability. For example, how might these students respond to questions about boredom or resentment which might be present as they experience years of repetitive math instruction? What are their attitudes toward acquiring labels like "geek" or "brainiac" if they display the everyday-achievement-oriented behaviors proposed above? What coping mechanisms do they choose, if such labels are seen as problematic? If there is a necessary and sufficient level of math enrichment needed to allay negative attitudes, how might that be documented through survey questions about these factors? Might it be possible to empirically identify the *faux* confidence that was proposed above, and is this a detrimental attitude if it is present? These are all questions which could be answered through further research.

A final limitation is embedded in conclusions which may may be drawn from the current research and impact on curricular or policy decisions. While this research seemed to show the high ability students fare reasonably well despite not being accelerated, and that they appeared to maintain more favorable attitudes toward math than their less able peers, it is not accurate to interpret the results as evidence against acceleration. Acceleration should still be one of the many options provided for gifted math students. The research (as reviewed above) on acceleration in its many forms has found it to be positive in nearly every way, and there is anecdotal evidence that math accelerative options create very positive results, in both achievement and attitude, for elementary students of high ability. It is highly probable that the positives documented for pre-school/primary math gifted students (Robinson et al, 1996/1997) and for secondary students (Benbow and Ajamund, 1990; Benbow, Lubinski and Suchy, 1996, Lubinski et al., 2001) would accrue to the later elementary aged students, as well. The challenge now is to empirically document these positive outcomes for this age group.

Summary

This research study was undertaken to determine if there were negative effects on the attitude of bright elementary math students, if they were not allowed to accelerate in their study of math. The concern that mathematically gifted student might lose interest and motivation without acceleration in math during the preadolescent years was not generally supported by this research. In fact, in most ways, the more able the student, the more positive the attitudinal factors. There was an incremental increase in anxiety, and there were incremental decreases: Perception of the teacher's view, enthusiasm for problem solving, success in math as socially desirable, and usefulness of math. However, these declines were less severe for the more mathematically able students than for the general population, and were off-set by consistent increases for the stronger students in other attitudinal factors--factors which declined to a significant degree in the general population. Many of the strong students would be enrolled in prealgebra as seventh graders, which may explain why that age is the point at which the brightest students experienced a jump in their perceptions of their teachers' attitudes toward them, and their view of math as useful. It could be that there would be no decline for the strong students, if they were offered accelerative options earlier in their school careers. The current research can neither support nor refute that possibility, though anecdotal evidence from previous research seems to suggest this, and the up tick in scores by students in the current research when they first experience acceleration is compelling evidence.

Though these results are encouraging for the cultivation of math talent, they must be interpreted with some caution. The sample was drawn from a single affluent district which provided enrichment options throughout elementary school, and which was, *de facto,* an enriched environment due to the generally high level of math achievement. High ability math students are not likely to be isolated from other high ability peers, nor to encounter more than occasional mediocre teaching. In other settings, such as inner cities, rural, or even less affluent suburbs, these environmental factors might bode differently for the young mathematically gifted student. Replication studies in these settings would be warranted.

Another caution is that the findings in this study should not be used to argue against acceleration of young, precocious students. The arguments against acceleration are shown to be chimeras in the research reviewed above. Further, the current research gives indirect support to the proposition that failing to acknowledge math ability publicly (acceleration would be only one way of public acknowledgement) may create a classroom climate of lower expectations for all, and contribute to a general decline in enthusiasm for math. Because acceleration leads to positive academic outcomes, and may lead to other positive outcomes in the psychosocial environment, it should be practiced more widely than it is in elementary and middle schools.

The gender differences were consistent with previously reported studies with regard to the persistent and troubling finding that boys continue to view math through a more stereotypical lens than their female peers. Boys were also significantly more confident and had less anxiety than girls of the same age, and held the more favorable view in most attitudes even when the difference did not reach levels of significance. This trend reversed, however, for the most able girls at the seventh grade. Among students at the highest ability, there was an insignificant gender discrepancy on most attitude factors until the seventh grade when girls reported more favorable views than boys. The congruence of attitudes for the highest ability students is at variance with previous research and is not easily explained; it may only reflect ideosyncracies of the small sample.

A serendipitous outcome of the current research is the validity of using the adapted Fennema Sherman to study math attitudes with a much younger cohort. This finding can provide a way to study the attitude issue among preadolescent and early middle school gifted students, a group which has received little attention in the research. A validation study with the general population of preadolescent students would be a worthy effort for further research.

Endnotes

1 The School and College Ability Tests (SCAT), at the time of the reported projects, were published by Educational Testing Service (ETS, Princeton, NJ). In 1996, however, ETS sold the publication rights to the Institute for Academically Advanced Youth (IAAY) at Johns Hopkins University. Currently, IAAY will make the SCAT available for a nominal licensing fee to districts wishing to replicate the SMPY model. Eventual publication and renorming studies are being explored (Carol Mills, personal communication, Nov., 1996).

2 In the study of the K-2 populations, Robinson, Abbott, Berninger, Busse and Mukhopadhyay (1997) describe as "disheartening" results showing the early and persistent appearance of gender differences within this highly advanced group. Even with significant outreach effort to parents of girls prior to beginning the study, the initial testing identified boys at a ratio of nearly 2:1. Post study, though both boys and girls in both groups showed gains, and the treatment group showed larger gains for both genders than the control group, the girls' gains in mathematics were insufficient to catch up to the boys after two years. This finding was specific to math, as neither gender was favored by the gains, in both groups, in verbal skills (Waxman, Robinson and Mukhopadhyay, 1996).

3 This could result in an accelerated pace of instruction, as well, especially in situations where students test out of preliminary material and are then permitted to self-pace themselves through the content. Students in the SMPY summer math programs cover a year or more of "standard" curriculum by condensing the typical 150 instructional hours to approximately 75 hours compacted into three weeks. Northwestern University's Center for Talent Development , which sponsors the Electronic Programs

for Gifted Youth e-mail correspondence courses, reports that students--most of whom are seventh or eighth graders--typically complete the Algebra I and II sequence (two years of study in high school) in 6 to 8 months, working at their own pace ("Letter Links," 2000) . Southern and Jones (1991) note that besides the administrative provisions vs. curricular-pacing-or-demands dichotomy, the literature also distinguishes between degrees of acceleration , ranging from modest (e.g., testing out of a chapter or single unit of material) to radical (e.g., two or more years of grade skipping, or fast paced summer programs like the SMPY model).

4 A concerted attempt was made to find a copy of the district's policy on acceleration in math. A request was made of the district's Curriculum Director, Math Coordinator, and the Coordinator of Gifted Services (going back for a period of 10 years at the time of the inquiry). None of them was able to actually produce such a written policy, though all agreed that it was, indeed, the policy not to accelerate. Further random inquiries to long-time elementary school principals, middle school principals, and math teachers at several levels produced similar responses. There were two rationale given for this longstanding policy-of-practice. First, it was firmly believed that research showed that students who moved through math too quickly "burned out" by later high school, and dropped out of advanced math courses. Second, there was significant concern that accelerating at the elementary level would provide a weak foundation for the students who would then be taking algebra as sixth or seventh graders ("elementary teachers don't understand advanced math well enough to teach the prerequisites"). Furthermore, there was a high level of concern that the students who were accelerated would become a *de facto* ability group in the middle school, due to scheduling issues. This, it was believed, could result in lower self confidence of students who were less able at math, and to significant parental pressure to have students placed in the "smart" group. Requests for

evidence or research supporting each of these rationale never yielded any confirming results with the exception of the parental pressure issue, which did have some anecdotal support.

5 The SCAT-Q for the class of 2005 was administered during the last two weeks of January of their fourth grade year. This placed them on the cusp between the fall and spring norms. The 2006 cohort was tested in April; therefore, for the sake of consistency, the Spring, seventh grade norms are used for both cohorts. This yields a conservative estimate of the numbers of high ability students. Had the Fall, seventh grade norms been used with the 2005 cohort, there would have been 77 students instead of 41 students scoring > 65%ile, or a total of 118 students instead of 83 for the two years.

6 Because raw score data is not available for the original F-S norming group, it is not possible to compute alternate split half reliabilities, or to figure standard deviations for different subsets of the scales in order to conduct t-tests or other comparative tests of means. However, the MANOVAs conducted in the current research to compare the means of various groups of interest, revealed that a difference of approximately 1 point in the means of the raw scores for the six-item positively phrased subscales produced a $p < .10$ (two tailed). Using this somewhat rough estimate as a gauge, a comparison of the means for the six negative and six positively phrased items, as reported in the F-S manual (1976), was conducted. The means of seven of the nine scales were less than 1 raw score point apart, suggesting that the positively and negatively phrased items contributed approximately equally to the subscale score. The two subscales in which the raw score means differed by more than 1 point were Anxiety and Effectance Motivation. In each case, the negatively phrased items contributed somewhat more heavily to a positive (i.e., favorable attitude toward math) result. Thus, for these two subscales, the

results obtained by using the adapted (positive) scale might be considered slightly skewed toward an unfavorable attitude. For the purposes of the current study, the comparisons between groups should not be impacted, as all groups participated with the same adapted scale and are being compared to each other, not to an external norm.

7 The finding of a substantial increase in the high-high girls' Father score, even surpassing the boys at the seventh grade, is congruent with findings reported by Miles (1985), in a qualitative study of 44 top women executives. One of the distinguishing characteristics of these women, in contrast to other women who did not move beyond middle-management level, was a father whom they greatly admired, and who provided them with both encouragement in their pursuits, and "protection" from the pressures to "become a young lady" when they reached puberty.

APPENDIX

Middle School Mathematics Attitude Assessment Report

Middle School Mathematics Attitude Assessment Report 1995-97

Introduction

During the 1996-97 school year, the School Improvement Team continued with its School Improvement (North Central Accreditation) academic goal of "improved achievement in mathematics". The team is monitoring the differences from year to year between male and female performance for each grade level. Student performance will be collected and monitored as part of the ongoing NCA process (See Middle School's NCA Report).

In addition to mathematics performance, the school continues to monitor all students' attitudes toward mathematics through the use of an adapted version of the Fennema-Sherman Mathematics Attitude Scales. All students in grades six through eight participate in the survey (See Exhibit A in the Appendix). The survey is administered by the staff and scored by the County Testing/Research Department. Additionally, open-ended questions are administered and reviewed by staff.

This report offers an overall view of the students' attitudes toward mathematics. Presented by grade level and gender, the results show patterns of year-to-year change among students who have participated in the survey for the past three years.

Survey Description

The modified Fennema-Sherman Mathematics Attitude Survey is divided into nine scales:

1) **Mother**, 2) **Father**, 3) **Teacher** (student's perceptions of teachers' attitudes toward them as learners of mathematics), 4) **Confidence** (willingness to attempt to learn mathematics, 5) **Success** (degree to which students attribute mathematics achievement to positive/negative consequences, e.g., "It was just luck!"), 6) **Male Domain** (mathematics is perceived to be appropriate to males), 7) **Usefulness** (measures the student's belief about the necessity of mathematics for current and future endeavors), 8) **Math Anxiety** (measures apprehension toward mathematics), and 9) **Effectance Motivation** (a persistence in wanting to engage in problem solving activities).

For specific examples of the nine scales use the sample survey found in Exhibit A of the Appendix. The following are the specific item numbers:

	Items #		Items #
Mother	= 2, 11, 20, 29, 38, 47	Male Domain	= 6, 15, 24, 33, 42, 51
Father	= 3, 12, 21, 30, 39, 48	Usefulness	= 7, 16, 25, 34, 43, 52
Teacher	= 5, 14, 23, 32, 41, 50	Anxiety	= 8, 17, 26, 35, 44, 53
Confidence	= 1, 10, 19, 28, 37, 46	Effectance Motivation	= 9, 18, 27, 36, 45, 54
Success	= 4, 13, 22, 31, 40, 49		

There are a total of six items for each scale. The items are positive statements on which students are asked to rate themselves from 1 to 5 (1 = strongly disagree to 5 = strongly agree), as to the extent to which they disagree or agree with the statements. The total range score for the attitudinal scale is 6 through 30.

Chart 1

Score		6-Item Rating Scale
30	=	5 (Strongly agree)
24 - 30	=	4 or 5
18 - 24	=	3 or 4
12 - 18	=	2 or 3
6 - 12	=	1 or 2

Participation Rate

Student participation rates appear below in Table 1.

Table 1

Participation Rates by Grade

Grade	Enrollment			Number Tested			Percent		
	'94/5	'95/6	'96/7	'94/5	'95/6	'96/7	'94/5	'95/6	'96/97
6	188	213	231	174	201	227	92.6	94.4	98.3
7	230	191	222	213	167	212	92.6	87.4	95.5
8	223	234	197	216	213	191	96.9	91.0	97.0
Total	641	638	650	603	581	630	94.1	91.1	96.9

Survey Results

The following series of tables report the scores for each scale. The use of a **median** score (fifty percent of population are above this point and fifty percent below) was selected for the overall scores because it best illustrates the midpoint for the total group. For grade level, reporting the **mean** score (arithmetic average) is utilized. However, it is interesting that the **mode** (the most frequent score) for many scales was extremely high. Each of these is further illustrated in Exhibits B, C, and D.

A composite summary of student results are reported by the median score, accounting for the use of whole numbers on Table 2 below.

Table 2

SUMMARY of MEDIAN SCORES*
ALL 3 GRADES

SCALE	1995	1996	1997	CHANGE** (1997-1995)
Confidence	24	24	25	+1
Mother	26	28	28	+2
Father	26	27	27	+1
Success	26	27	28	+2
Teacher	21	22	23	+2
Male Domain	28	30	29	+1
Usefulness	26	26	26	=
Math Anxiety	21	21	21	=
Effectance Motivation	19	19	20	+1

1994/5 N = 603 (Males=303) (Females=300)
1995/6 N= 581 (Males=293) (Females=288)
1996/7 N= 630 (Males=302) (Females=328)

* Based on total score of 30.
** For the purpose of interpretation of this survey, significance is defined as a ± 2 points (6.7%)

Table 3

Composite of Overall Mean Scores*

Scale	Grade 6			Grade 7			Grade 8			All Grades		
	'95	'96	'97	'95	'96	'97	'95	'96	'97	'95	'96	'97
Confi-dence	23.6	24.1	25.2	23.8	23.1	23.0	23.1	23.8	23.6	23.5	23.7	23.9
Mother	25.3	26.6	26.3	25.3	26.2	26.6	25.4	25.9	26.2	25.3	26.3	26.4
Father	24.8	26.2	26.1	25.0	25.7	26.0	25.4	25.5	25.9	25.1	25.8	26.0
Success	26.3	26.5	27.2	24.9	25.7	25.7	24.8	25.2	25.0	25.3	25.8	26.0
Teacher	21.8	23.3	23.2	19.8	20.4	21.2	21.9	21.8	22.4	21.1	21.9	22.3
Male Domain	27.2	27.7	27.2	26.5	27.3	27.4	26.3	27.1	26.6	26.6	27.3	27.1
Useful	24.6	25.6	25.9	24.7	24.4	24.9	24.8	24.5	24.0	24.7	24.9	25.0
Math Anxiety	20.5	21.0	22.3	21.0	19.8	19.8	20.5	20.8	20.9	20.7	20.6	21.1
Effect. Motiva-tion	19.9	20.4	22.0	19.3	18.7	18.6	18.4	18.5	18.5	19.2	19.2	19.8

Over the three-year period, the overall student ratings report a consistently similar pattern with respect to each scale and across each grade level.

The one exception is at grade six where Effectance Motivation and Math Anxiety showed a significant increase.

* Based on total score of 30.
For the purpose of interpretation of this survey, significance is defined as a ± 2 points (6.7%)

Table 4

Summary of Overall Mean Scores*

Grade 6

Scale:	Male '95	Male '96	Male '97	Female '95	Female '96	Female '97	Total '95	Total '96	Total '97
Confidence	24.6	24.7	25.9	22.7	23.4	24.5	23.6	24.1	25.2
Mother	25.3	27.1	26.5	25.2	26.2	26.1	25.3	26.6	26.3
Father	24.8	26.4	26.3	24.9	26.0	25.9	24.8	26.2	26.1
Success	26.0	26.4	27.4	26.6	26.5	26.9	26.3	26.5	27.2
Teacher	21.4	22.9	23.3	22.2	23.5	23.2	21.8	23.3	23.2
Male Domain	25.7	26.7	26.0	28.5	28.6	28.6	27.2	27.8	27.2
Usefulness	24.4	25.4	26.2	24.7	25.8	25.5	24.6	25.6	25.9
Math Anxiety	21.6	21.7	23.2	19.6	20.3	21.4	20.5	21.0	22.3
Effectance Motivation	20.3	20.1	22.0	19.5	20.7	22.1	19.9	20.4	22.0

For grade six, when considering male and females together, the overall pattern of attitudinal rankings indicate significant increases in Effectance Motivation and Math Anxiety. All other categories are reporting slight to moderate increases in ratings with only the rankings for Teacher and Confidence approaching the level of significance.

When analyzed separately, males and females display the same pattern. The only exception being the males who show a slightly stronger increase in their ratings of Teacher than females over the three-year period.

* Based on total score of 30.
For the purpose of interpretation of this survey, significance is defined as a ± 2 points (6.7%)

Table 5

Summary of Overall Mean Scores*

Grade 7

Scale	Male			Female			Total		
	'95	'96	'97	'95	'96	'97	'95	'96	'97
Confidence	24.8	24.6	24.2	22.6	21.8	21.9	23.8	23.1	23.0
Mother	25.5	25.9	27.1	25.2	26.6	26.1	25.3	26.3	26.6
Father	24.9	25.5	26.5	25.2	25.9	25.5	25.0	25.7	26.0
Success	25.4	25.1	26.2	24.2	26.1	25.3	24.9	25.7	25.7
Teacher	20.2	21.2	22.0	19.2	19.6	20.6	19.8	20.4	21.2
Male Domain	24.5	26.0	26.4	29.0	28.4	28.3	26.5	27.3	27.4
Usefulness	25.0	24.6	25.0	24.4	24.2	24.8	24.7	24.4	24.9
Math Anxiety	21.7	21.3	20.8	20.1	18.5	19.2	21.0	19.8	19.8
Effectance Motivation	19.8	19.5	19.3	18.7	17.9	18.0	19.3	18.7	18.6

Over the three-year period, at grade seven, when considering both males and females, there appears to be no significant change in any of the ratings of the nine scales. The ratings appear constant with only slight variations from year to year.

The only changes of note were with male students in the areas of Teacher and Male Domain. Over the three year period the increase in these two scales approached the significance level with increases of 1.8 and 1.9 respectively.

* Based on total score of 30.
For the purpose of interpretation of this survey, significance is defined as a ± 2 points (6.7%)

Table 6

Summary of Overall Mean Scores*

Grade 8

Scale	Male			Female			Total		
	'95	'96	'97	'95	'96	'97	'95	'96	'97
Confidence	23.7	24.4	24.0	22.6	23.2	23.2	23.1	23.8	23.6
Mother	25.8	25.8	25.8	25.1	26.0	26.6	25.4	25.9	26.2
Father	26.3	25.1	25.5	25.0	26.0	26.1	25.4	25.2	25.9
Success	24.5	25.0	24.2	25.1	25.4	25.6	24.8	25.2	25.0
Teacher	21.9	22.6	21.8	22.1	20.9	22.8	22.0	21.8	22.4
Male Domain	23.8	25.5	24.6	28.5	29.0	28.2	26.3	27.1	26.6
Usefulness	25.0	24.9	23.5	24.5	24.1	24.4	24.8	24.5	24.0
Math Anxiety	21.9	21.8	21.9	19.5	19.7	20.1	20.5	20.8	20.9
Effectance Motivation	18.4	19.2	19.3	18.4	17.7	17.8	18.4	18.5	18.5

For grade eight, with the possible exception of the females' ratings of Mother, the three-year ratings for all nine scales reflect a remarkably consistent pattern of responses.

* Based on total score of 30.
For the purpose of interpretation of this survey, significance is defined as a ± 2 points (6.7%)

8

Table 7

Summary of Overall Mean Scores*

All Grades Combined

Scale	Male '95	Male '96	Male '97	Female '95	Female '96	Female '97	Total '95	Total '96	Total '97
Confidence	24.4	24.6	24.8	22.6	22.9	23.2	23.5	23.7	23.9
Mother	25.5	26.3	26.5	25.2	26.2	26.3	25.3	26.3	26.4
Father	25.7	25.6	26.2	24.9	26.0	25.8	25.1	25.8	26.0
Success	25.9	25.5	26.1	25.3	26.0	25.9	25.3	25.8	26.0
Teacher	21.1	22.4	22.5	21.2	21.5	22.2	21.2	21.9	22.3
Male Domain	24.6	26.0	25.7	28.7	28.7	28.4	26.6	27.3	27.1
Usefulness	24.9	25.0	25.1	24.5	24.7	24.9	24.7	24.9	25.0
Math Anxiety	21.6	21.6	22.0	19.7	19.6	20.2	20.7	20.6	21.1
Effectance Motivation	19.5	19.6	20.4	18.8	18.8	19.3	19.2	19.2	19.8

As expected, when viewing the mean scores for all grades on all nine scales reflecting students perceptions about mathematics, there is no change which registers at the level of significance. Changes from 1995 to 1996 to 1997 have been minimal. The rating of mathematics as a Male Domain has been the highest rated area by both genders with the female perception ratings higher than the males.

* Based on total score of 30.
For the purpose of interpretation of this survey, significance is defined as a ± 2 points (6.7%)

CORRELATION OF MATH GRADE WITH SURVEY RESULTS
April, 1997

Student responses on the Fennema-Sherman rating scales were matched with the most recent letter grade students received in their mathematics class. The Pearson Test of Significance was used to determine the degree of statistical significance among grade levels and the group as a whole. Summary results are reported for difference at/below the .05 level of significance. When correlating the mathematics grades with the attitude scales the following correlations are noted:

1) As expected, students at all three grade levels who earned A's displayed significantly more Confidence in their mathematics ability than students achieving B's or lower.

2) On the scale Mother, at two grade levels there was a significant difference reported.

   Seventh grade students, achieving B's reported Mother's influence less positively than the other achievement levels.

   At the eighth grade level, C-D-E students, likewise, report Mother's influence less positively than the other achievement levels.

3) On the Success scale (which attributes mathematics achievement to positive or negative consequences), eighth grade students earning A's are significantly more positive in their responses. Students earning C-D-E's rated Success significantly more negatively in their responses.

4) On the scale Male Domain (which indicates mathematics to be appropriate to males), at the eighth grade level a significant number of students earning A's perceive mathematics to be a male domain while students earning C-D-E's significantly reported that it was not.

5) Although not significant at specific grade levels, when analyzed by the total population of students, the perception of Usefulness of mathematics by students achieving A's was significantly higher than the students achieving B's and lower in mathematics.

6) At every grade level, students earning C-D-E's indicated a significantly high level of Math Anxiety. This is also true for seventh grade students earning B's.

7) Sixth and seventh students earning C-D-E's on the Effectance Motivation scale report significantly lower ratings than A and B students. Effectance Motivation reflects one's engagement in problem solving activities.

EXHIBIT A

Mathematics Attitude Scales

On the following pages are a series of statements. There are no correct answers for these statements. They have been set up in a way which permits you to indicate the extent to which you agree or disagree with the ideas expresses. Suppose the statement is:

Example 1. I like mathematics.

As you read the statement, you will know whether you agree or disagree. If you strongly agree, blacken circle A opposite the number on your answer sheet. If you agree but with reservations, that is, you do not fully agree, blacken circle B. If you disagree with the idea, indicate the extent to which you disagree by blackening circle D for disagree or circle E if you strongly disagree. But if you neither agree nor disagree, that is, you are not certain, blacken circle C for undecided. Also, if you cannot answer a question, blacken circle C. Now mark your answer sheet. Do the same for example No. 2.

Example 2. Math is very interesting to me.

Do not spend much time with any statement, **but be sure to answer every statement.** Work fast but carefully.

There are no "right" or "wrong" answers. The only correct responses are those that are true **for you.** Whenever possible, let the things that have happened to you help you take a choice. Do not mark on the booklet.

THIS INVENTORY IS BEING USED FOR RESEARCH PURPOSES ONLY AND NO ONE WILL KNOW WHAT YOUR RESPONSES ARE.

C 1. Generally I feel secure about attempting mathematics.
M 2. My mother thinks I can do well in mathematics.
F 3. My father thinks that mathematics is one of the most important subjects I have studied.
S 4. It would make me happy to be recognized as an excellent student in mathematics.
T 5. My teachers have encouraged me to study more mathematics.
MD 6. Females are as good as males in geometry.
U 7. I'll need mathematics for my future work.
A 8. Math doesn't scare me at all.
EM 9. I like math puzzles.
C 10. I am sure I could do advanced work in mathematics.
M 11. My mother thinks I could be good in math.
F 12. My father has strongly encouraged me to do well in mathematics.
S 13. I'd be proud to be the outstanding student in math.
T 14. My teachers think I can do well in mathematics.
MD 15. Studying mathematics is just as appropriate for women as for men.

U 16. I study mathematics because I know how useful it is.
F 17. It wouldn't bother me at all to take more math courses.
EM 18. Mathematics is enjoyable and stimulating to me.
C 19. I am sure that I can learn mathematics.
M 20. My mother has always been interested in my progress in mathematics.
F 21. My father has always been interested in my progress in mathematics.
S 22. I'd be happy to get top grades in mathematics.
T 23. Math teachers have made me feel I have the ability to go on in mathematics.
MD 24. I would trust a woman just as much as I would trust a man to figure out important calculations.
U 25. Knowing mathematics will help me earn a living.
F 26. I usually don't worry about being able to solve math problems.
EM 27. When a math problem arises that I can't immediately solve, I stick with it until I have the solution.
C 28. I think I could handle more difficult mathematics.
M 29. My mother has strongly encouraged me to do well in mathematics.
F 30. My father thinks I'll need mathematics for what I want to do after I graduate from high school.
S 31. It would be really great to win a prize in mathematics.
T 32. My math teachers would encourage me to take all the math I can.
MD 33. Girls can do just as well as boys in mathematics.
U 34. Mathematics is a worthwhile and necessary subject.
F 35. I almost never have gotten shook up during a math test.
EM 36. Once I start trying to work on a math puzzle, I find it hard to stop.
C 37. I can get good grades in mathematics.
M 38. My mother thinks that mathematics is one of the most important subjects I have studied.
F 39. My father thinks I could do well in mathematics.
S 40. I would be pleased to be first in a mathematics competition.
T 41. My math teachers have been interested in my progress in mathematics.
MD 42. Males are naturally better than females in mathematics.
U 43. I'll need a firm mastery of mathematics for my future work.
F 44. I am usually at ease in math classes.
EM 45. When a question is left unanswered in math class, I continue to think about it afterward.
C 46. I have a lot of self-confidence when it comes to math.
M 47. My mother thinks I'll need mathematics for what I want to do after I graduate from high school.
F 48. My father thinks I could be good in math.
S 49. It would be great to be regarded as smart in mathematics.
T 50. I could talk to my math teachers about a career which uses math.
MD 51. Women certainly are logical enough to do well in mathematics.
U 52. I will use mathematics in many ways as an adult.
F 53. I usually have been at ease in math classes.
EM 54. I am challenged by math problems I can't understand immediately.

RESOURCES

Resources

Ablard, K. E. & Tissot, S. L. (1998). Young students' readiness for advanced math: Precocious abstract reasoning. *Journal for the Education of the Gifted, 21* (2), 206-223.

American Association for Gifted Children (1978). *On being gifted.* New York: Walker.

American Association of University Women (1990). *Shortchanging girls, shortchanging America: Full data report.* Washington, DC: American Association of University Women.

American Association of University Women (1992). *How schools shortchange girls: A study of the major findings on girls and education, AAUW report.* Washington, DC: The AAUW Educational Foundation and National Educational Association.

Anderson, L. W. & Bourke, S. F. (2000). *Assessing affective characteristics in the schools. 2nd. ed.* Mahwah, New Jersey: Lawrence Erlbaum Associates.

Arnold, K., Noble, K. D. & Subotnik, R. F. (1996). Perspectives on female talent development. In K. Arnold, K. D. Noble, & R. F. Subotnik. *Remarkable women: Perspectives on female talent development* (pp. 1-23). Cresskill, N. J.: Hampton Press, Inc.

Assouline, S. G. & Lupkowski-Shoplik, A. (1997). Talent searches: A model for the discovery and development of academic talent. In N. Colangelo & G. Davis (Eds.), *Handbook of gifted education, 2nd. ed.* (pp. 170-179). Boston: Allyn and Bacon.

Becker, B. J. & Hedges, L. V. (1988). The effects of selection and variability in studies of gender differences. *Behavioral and Brain Sciences, 11* (2), 183-185.

Belenky, M. F., Clinchy, B. M., Goldberger, N. R. & Tarule, J. M. (1986). *Women's ways of knowing: The development of self, voice and mind.* New York, Basic Books.

Benbow, C. P. (1986). SMPY's model for teaching mathematically precocious students. In J. S. Renzulli (Ed.), *Systems and models for developing programs for the gifted and talented* (pp. 2-25). Mansfield Center, CT: Creative Learning Press.

Benbow, C. (1988). Sex differences in mathematical reasoning ability in intellectually talented preadolescents: Their nature, effects, and possible causes. *Behavioral and Brain Sciences, 11* (2), 169-232.

Benbow, C. & Arjmand, O. (1990). Predictors of high academic achievement in mathematics and science by mathematically talented studetns: A longitudinal study. *Journal of Educational Psychology, 82* (3), 430-441.

Benbow , C. & Lubinski, D. (1997). Intellectually talented children: How can we best meet their needs? In N. Colangelo & G. Davis (Eds.), *Handbook of gifted education, 2nd. ed.* (pp. 155-169). Boston: Allyn and Bacon.

Benbow, C., Lubinski, D., & Suchy, B. (1996). The impact of SMPY's educational programs from the perspective of the participant. In C. P. Benbow & D. Lupinski (Eds.), *Psychometirc and social issues: Intellectual talent* (pp. 266-300). Baltimore, MD: Johns Hopkins University Press.

Benbow, C. & Stanley, J. (1980). Sex differences in mathematical ability: Fact or artifact? *Science 210.* 1262-1264.

Benbow, C. & Stanley, J. (1981). Mathematical ability: Is sex a factor? *Science, 212.* 114-121.

Boatman, T. A., Davis, K. G. & Benbow, C. P. (1995). Best practices in gifted education. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology III* (pp. 1083-1095). Washington DC: National Association of School Psychologists.

Brody, L. E. & Blackburn, C. C. (1996). Nurturing exceptional talent: SET as a legacy of SMPY. In C. P. Benbow & D. Lupinski (Eds.), *Psychometirc and social issues:Intellectual talent.* (pp. 246-265). Baltimore, MD: Johns Hopkins University Press.

Brown, L. M. & Gilligan, C. (1992). *Meeting at the crossroads.* New York: Random House.

Callahan, C. (1991). An update on gifted females. *Journal for the Education of the Gifted, 14* (3), 284-311.

Charlton, J.C., Marolf, D. M. & Stanley, J. C. (1994). Follow-up insights on rapid educational acceleration. *Roeper Review, 17*, 123-130.

Cogan, L. & Schmidt, W. H. (1999). Middle school math reform. *Middle Matters, 8* (1), 2-3.

Cohen, J. (1991). A power primer. *Psychological Bulletin, 112* (1), 155-159.

Comprehensive test of basic skills (CTBS), fourth edition. (1991). Monterey, California: CTB.

Cossey, R. (1999). Are California's math standards up to the challenge? *Phi Delta Kappan, 80* (6), 441-443.

Cramer, R. H. (1989). Attitudes of gifted boys and girls towards math: A qualitative study. *Roeper Review, 11* (3), 128-130.

Daniel, N. & Cox, J. (1988). *Flexible pacing for able learners.* Reston, VA: Council for Exceptional Children.

Delcourt, M. A., Loyd, B. H., Cornell, D. G. & Goldgerg, M. D. (1994). *Evaluation of the effects of programming arrangements on student learning outcomes.* Storrs, CT: NRC/GT, University of Connecticut.

Delisle, J. R. (1987). *Gifted kids speak out.* Minneapolis, MN: FreeSpirit Press.

Eccles, J. (1985). Why doesn't Jane run? Sex differences in educational and occupational patterns. In F. D. Horowitz & M. O'Brien, (Eds.) *The gifted and talented: Developmental perspectives.* (pp. 251-300). Washington, D.C.: American Psychological Association.

Eccles, J. S., Wigfield, A., Harold, R. & Blumenfeld, P. (1993). Age and gender differences in children's achievement self-perceptions during the elementary school years. *Child Development, 64*, 830-847.

Ewers, C. A. & Wood, N. L. (1993). Sex and ability differences in children's math self-efficacy and prediction accuracy. *Learning and individual differences, 5* (3), 259-267.

Feldhusen, J., Proctor, T. B. & Black, K. N. (1986). Guidelines for grade advancement of precocious children. *Roeper Review, 9* (1), 25-27.

Fennema, E. & Hart, L. (1994). Gender and the JRME. *The Journal for Research in Mathematics Education, 25* (6), 648-659.

Fennema, E. & Sherman, J. (1976). Fennema-Sherman Mathematics Attitudes Scales. Princeton, NJ; Educational Testing Service-ETS Test Collection.

Fennema, E. & Peterson, P.L. (1985). Autonomous learning behavior: A possible explanation of gender-related differences in mathematics. In L.C. Wilkinson & C. B. Marrett (Eds.), *Gender-related differences in classroom interactions* (pp. 17-35). New York: Academic Press.

Fox, L. H. (1976). Sex differences in mathematical precocity: Bridging the gap. In D. P. Keating (Ed.), *Intellectual talent: Research and development* (pp. 183-214). Baltimore: Johns Hopkins University Press.

Fox, L H. (1981). Identification of the academically gifted. *American Psychologist, 36* (10), 1103-1111.

Frome, P. & Eccles, J. S. (1998). Parents' influence on children's achievement-related perceptions. *Journal of Personality and Social Psychology, 74* (2), 435-452.

Fuligni, A. J., Eccles, J. S., & Barber, B. L. (1995). The long-term effects of seventh-grade ability grouping in mathematics. *Journal of Early Adolescence, 15* (1), 58-89.

Galbraith, J. & Delisle, J. (1996). *The gifted kids survival guide: A teen handbook. (revised, expanded and updated edition).* Minneapolis, MN: Free Spirit Press.

George, W.C., Cohn, S. J., & Stanley, J. C. (Eds.). (1979). *Educating the gifted: Acceleration and enrichment.* Baltimore: Johns Hopkins University Press.

George, O., Stevenson, C., Thomason, J., & Beane, J. (1992). *The middle school--and beyond.* Alexandria, VA: Association for Supervision and Curriculum Development.

Giarelli, E. (1995). Gender issues: Book reviews. *Educational Studies, 26,* 36-41.

Gilligan, C. (1977). In a different voice: Women's conceptions of self and morality. *Harvard Educational Review, 47,* 481-517.

Gilligan, C. (1982). *In a different voice: Psychological theory and women's development.* Cambridge, MA: Harvard University Press.

Higham, S. J. & Buescher, T. M. (1987). What young gifted adolescents understand about feeling "different". In T. Buescher (Ed.), *Understanding gifted and talented adolescents: A resource guide for counselors, educators, and parents.* (pp. 77-94). Evanston, IL: Center for Talent Development, Northwestern Univ.

Hill, J. P. & Lynch, M. E. (1983). The intensification of gender-related role expectations during early adolescence. In J. Brooks-Gunn & A. C. Petersen (Eds.), *Girls* *at puberty.* (pp. 201-228). New York: Plenum.

Hughes, M. E. (1984). *A study of mathematically gifted girls, their math achievement, sex role stereotyping, and attitudes toward math related careers.* Unpublished doctoral dissertation, Columbia University Teachers College, N.Y.

Hunt, B. G. (1996). *The effect of homogeneous and heterogeneous grouping of gifted sixth grade students on mathematics achievement and attitude.* Unpublished doctoral dissertation: University of Houston, Houston, TX.

Hunt, B. (1996/97). The effect on mathematics achievement and attitude of homogeneous and heterogeneous grouping of gifted sixth-grade students. *The Journal of Secondary Gifted Education, 7* (2), 65-73.

Hyde, J.S. (1981). How large are cognitive gender differences? A meta-analysis. *American Psychologist, 36*, 892-901.

Hyde, J. S., Fennema, E. & Lamon, S. (1990). Gender differences in mathematics performance: A Meta-Analysis. *Psychological Bulletin, 107* (2), 139-153.

Hyde, J. S., Fennema, E., Ryan, M., Frost, L.A. & Hopp, C. (1990). Gender comparisons of mathematics attitudes and affect: A meta-analysis. *Psychology of Women Quarterly, 14,* 299-324.

Jensen, A. R. (1988). Sex differences in arithmetic computation and reasoning in prepubertal bays and girls. *Behavioral and Brain Sciences, 11* (2), 198-199.

Johnson, B. (2001). Toward a new classification of nonexperimental quantitative research. *Educational Researcher, 30* (2), 3-13.

Junge, M. E. & Dretzke, B. J. (1995). Mathematical self-efficacy gender differences in gifted/talented adolescents. *Gifted Child Quarterly 39* (1), 22-25.

Kauffman, J. M. & Hallaha, D. P. (Eds.). (1995). *The illusion of full inclusion : A comprehensive critique of a current special education bandwagon.* Austin, Tex. : Pro-Ed, 1995.

Kerr, B. (1994). *Smart girls two: A new psychology of girls, women and giftedness.* Dayton, OH: Ohio Psychology Press

Kerr, B. (1997). Developing talents in girls and young women. In N. Colangelo & G. Davis (Eds.), *Handbook of gifted education, 2nd. ed.* (pp. 483-497). Boston: Allyn and Bacon.

Kulik, J. A. & Kulik, C. L. C. (1984). Effects of accelerated instruction on students. *Review of Educational Research, 54* (3), 409-425.

Kulik, J. A. & Kulik, C. L. C. (1991). Ability grouping and gifted students. In N. Colangelo & G. Davis (Eds.), *Handbook of gifted education,* (pp. 178-196). Boston: Allyn and Bacon.

Kulik, J. A. & Kulik, C. L. C. (1997). Ability grouping. In N. Colangelo & G. Davis (Eds.), *Handbook of gifted education, 2nd. ed.* (pp. 230-242). Boston: Allyn and Bacon.

Letter Links (2000). Evanston, IL: Center for Talent Development/Midwest Talent Search.

Loveless, T. (1999 a). *The tracking wars: State reform meets school policy.* Washington, D.C.: Brookings Institution Press.

Loveless, T. (1999 b). Will tracking reform promote social equity? *Educational Leadership, 56* (7), 28-32.

Lupkowski-Shoplik, A. E. & Assouline, S. G. (1993). Identifying mathematically talented elementary students: Using the lower level of the SSAT. *Gifted Child Quarterly, 37* (3), 118-123.

Lubinski, D., Benbow, C.P. & Sanders, C. E. (1993). Reconceptualizing gender differences in achievement among the gifted. In K. A. Heller, F. J. Moenks, & A. H. Passow (Eds.), *International handbook of research and development of giftedness and talent.* . (pp. 693-707). Tarrytown, NY: Pergammon Press.

Lubinski, D., Webb, R. M., Morelock, M. J. & Benbow, C. P. (2001). Top 1 in 10,000: A 10-year follow-up of the profoundly gifted. *Journal of Applied Psychology, 86* (4), 718-729.

Maccoby, E. E. & Jacklin, C. N. (1974). *The psychology of sex differences.* London: Oxford University Press.

Meuller, D. J. (1986). *Measuring social attitudes: A handbook for researchers and practitioners.* New York: Teachers College Press.

Meyer, M R. & Koehler, M. S. (1990). Internal influences on gender differences in mathematics. In E. Fennema & G. Leder (Eds.), *Mathematics and gender.* (pp. 60-95). New York: Teachers College Press.

Miller, R., Mills, C. & Tangherlini, A. (1995). The Appalachia model mathematics program for gifted students. *Roeper Review, 18* (2), 138-141.

Miles, R. (1985). *Women and power.* London: Salem House.

Oakes, J. (1985). *Keeping track.* New Haven: Yale University Press.

Olszewski-Kubilius, P. (1998). Talent search. *Journal for Secondary Gifted Education, 9* (3), 106-113.

Ornstein, P. (1994). *School girls: Young women, self-esteem and the confidence gap.* New York: Doubleday.

Page, E B. & Keith, T. Z. (1996). The elephant in the classroom: Ability grouping and the gifted. In C.P Benbow & D. Lubinski (Eds.), *Intellectual talent: Psychometric and social issues.* (pp. 192-215). Baltimore, MD: Johns Hopkins University Press.

Pipher, M. (1994). *Reviving Ophelia: Saving the selves of adolescent girls.* New York: Putnam and Sons.

Reis, S. M. (1998). *Work left undone: Choices and compromises of talented females.* Mansfield Center, CT: Creative Learning Press.

Reis, S.M., Burns, D.J. & Renzulli, J. S. (1992). *Curriculum compacting: A complete guide to modifying the curriculum for high ability students.* Mansfield Center, CT.: Creative Learning Press.

Reis, S. & Callahan, C. (1989). Gifted females: They've come a long way--or have they? *Journal for the education of the gifted, 12* (2), 99-117.

Reis, S., Callahan, C. & Goldsmith, D. (1996). Attitudes of adolescent gifted girls and boys toward education, achievement and the future. In K. Arnold, K. D. Noble, & R. F. Subotnik. *Remarkable women: Perspectives on female talent development*. (pp. 209-225). Cresskill, N. J.: Hampton Press, Inc.

Renzulli, J. (1994). *Schools for talent development: A practical plan for total school imporvement.* Mansfield Center, CT: Creative Learning Press.

Reys, B., Robinson, E., Sconiers, S., & Mark, J. (1999). Mathematics curricula based on rigorous national standards. *Phi Delta Kappan, 80* (6), 454-456.

Robinson, N.M., Abbott, R.D., Berninger, V. W. & Busse, J. (1996). The structure of abilities in math-precocious young children: Gender similarites and differences. *Journal of Educational Psychology, 88,* 341-352.

Robinson, N.M., Abbott, R.D., Berninger, V.W. Busse, J. & Mukhopadhyay, S. (1997). Developmental changes in mathematically precocious young children: Longitudinal and gender effects. *Gifted Child Quarterly, 41* (4), 145-158.

Rogers, K. (1995, November). *Research-supported management practices in gifted education.* Paper presented at the annual conference of the National Association for Gifted Children. Tampa, FL.

Rogers, K. (1991). *The relationship of grouping practices to the education of the gifted and talented learner.* Storrs, CT: NRC/GT, University of Connecticut.

Rosenthal, R. (1988). Evaluating explanations of sex differences in mathematical reasoning scores. *Behavioral and Brain Sciences, 11* (2), 207-208.

Sadker, M. & Sadker, D. (1994). *Failing at fairness: How our schools cheat girls.* New York: Simon and Shuster.

Schmidt, W. H., McKnight, C, & Raizen, S. A. (1997). *A Splintered Vision: An Investigation of the U.S. Science and Mathematics Education--Executive Summary.* East Lansing, MI: U.S. National Research Center for the Third International Mathematics and Science Study, Michigan State University.

Sewall, G. (1993). How testbooks shortchange girls: A study of major findings on girls and education. *Society, 30,* 94-96.

Sells, L. (1973). *High school mathematics as the critical filter in the job market.* Berkely: University of California. (ERIC Document Reportduction Service No. ED 080 351).

Sheilds, C. J. (1992). "How schools shortchange girls"--a new report. *Curriculum Review, 31,* 3-5.

Slavin, R. (1987). Ability grouping and student achievement in elementary schools: A best-evidence synthesis. *Review of Educational Research, 57* 293-336.

Smith, A. E., Jussim, L., Eccles, J., VanNoy, M., Madon, S. & Palumbo, P. (1998). Self-fulfilling prophecies, perceptual biases and accuracy at the individual and group levels. *Journal of Experimental Social Psychology, 34,* 530-561.

Southern, W. T. & Jones, E. D. (1991). *The academic acceleration of gifted children.* New York: Teachers College Press.

Southern, W. T, Jones, E. D. & Stanley, J. C. (1993). Acceleration and enrichment: The context and development of program options. In K. A. Heller, F. J. Monks & H. A. Passow (Eds.) *Internation handbook of research and development of giftedness and talent* (pp. 387-410). Tarrytown, N Y: Pergamon Press.

Stanley, J.C. (1978). SMPY's DT-PI model: Diagnostic testing followed by prescriptive instruction. *Intellectually Talented Youth Bulletin, 4,* 7-8.

Stanley, J. C. (1991). An acedmic model for educating the mathematically talented. *Gifted Child Quarterly, 35,* 36-42.

Stanley, J. C. (1996). In the beginning: The study of mathematically precocious youth. In C. P. Benbow & D. Lupinski (Eds.) *Psychometirc and social issues: Intellectual talent* (p. 225-235). Baltimore, MD: Johns Hopkins University Press.

Stanley, J. C., Keating, D. P. & Fox, L. H. (Eds.) (1974). *Mathematical talent: Discovery, description, and development.* Baltimore, MD: Johns Hopkins University Press.

Sternberg, R. J. (1988). The male/female difference is there: Should we care? *Behavioral and Brain Sciences. 11* (2), 210-211.

Swiatek, M.A. & Benbow, C. P. (1991). A ten-year longitudinal follow-up of ability-matched accelerated and unaccelerated gifted students. *Journal of Educational Psychology, 83,* 528-538.

Swiatek, M. A. & Lupkowski-Shoplik, A. (2000). Gender differences in academic attitudes among gifted elementary school students. *Journal for the Education of the Gifted, 23 (*4), 360-377.

Terman, L. M. & Oden, M. H., *Genetic studies of genius,* Vol. IV. Stanford, CA: Stnaford University Press.

Tomlinson, C. (1995). *How to differentiate instruction in mixed abilty classrooms.* Alexandria, VA: Association for Supervision and Curriculum Development.

Tomlinson, C. (1999). *The differentiated classroom: Responding to the needs of all learners.* Alexandria, VA: Association for Supervision and Curriculum Development.

Van Tassel-Baska, J. (1981, December). Debate on acceleration vs. enrichment: The case for acceleration. Paper presented at the CEC-TAG Topical Conference on the Gifted and Talented, Orlando, FL.

VanTassel-Baska, J. (1992). Educational decision making on acceleration and grouping. *Gifted Child Quarterly, 36* (6), 68-71.

VanTassel-Baska, J. (1996). Contributions of the talent-search concept to gifted education. In C. P. Benbow & D. Lupinski, (Eds.) *Psychometirc and social issues: Intellectual talent* (p. 236-245). Baltimore, MD: Johns Hopkins University Press.

Waxman, B., Robinson, N., & Mukhopadhyay, S. (1996). *Teachers nurturing math-talented young children.* Storrs, CT: NRC/GT, University of Connecticut.

Webb, J. (1983). *Guiding the gifted child.* Dayton, OH: Ohio Psychology Press.

Westberg, K. L., Archambault, F. X, Dobyns, S. M. & Salvin, T. J. (1993). *An observational study of instrudtional and curricular practices used with gifted and talented students in regular classrooms.* Storrs, CT: NRC/GT, University of Connecticut.

Wigfield, A. & Eccles, J. S. (1992). The development of achievement task values: A theoretical analysis. *Developmental Review, 12*, 265-310.

Wigfield, A. & Eccles, J. S. (1994). Children's competence beliefs, achievement values and general self esteem: Change across elementary and middle school. *Journal of Early Adolescence, 14* (2), 107-138.

Wigfield, A., Eccles, J. S., MacIver, D., Reuman, D. A. & Midgley, C. (1991). Transitions during early adolescence: Changes in children's domain-specific self-perceptions and general self-esteem across the transition to junior high school. *Developmental Psychology*, *27*, (4), 552-565.

Wigfield, A., Eccles, J. S., Yoon, K. S., Harold, R., Arbreton, A, Freedman-Doan, C. & Blumenfeld, P. (1997). Change in children's competence beliefs and subjective task values across the elementary school years: A 3-year study. *Journal of Educational Psychology, 89* (3), 451-469.