Report Number FR 10-45

Exploration of FCAT Equating and Its Impact on Student, School, and District Developmental Scale Scores for 2003 through 2010

R. Gene Hoffman
Yvette M. Nemeth

Prepared for: Florida Department of Education
325 West Gaines Street
Tallahassee, FL 32399

Prepared under: Memorandum of Agreement, July 13, 2010

August 4, 2010

Human Resources Research Organization
66 Canal Center Plaza, Suite 700 • Alexandria, Virginia 22314
www.humrro.org • Phone: (703) 549-3611 - Fax: (703) 519-9661

Report Number FR 10-45

Exploration of FCAT Equating and Its Impact on Student, School, and District Developmental Scale Scores for 2003 through 2010

R. Gene Hoffman
Yvette M. Nemeth

Prepared for: **Florida Department of Education**
325 West Gaines Street
Tallahassee, FL 32399

Prepared under: **Memorandum of Agreement, July 13, 2010**

August 4, 2010

Human Resources Research Organization
66 Canal Center Plaza, Suite 700 • Alexandria, Virginia 22314
www.humrro.org • Phone: (703) 549-3611 - Fax: (703) 519-9661

EXPLORATION OF FCAT EQUATING AND ITS IMPACT ON STUDENT, SCHOOL, AND DISTRICT DEVELOPMENTAL SCALE SCORES FOR 2003 THROUGH 2010

Table of Contents

EXPLORATION OF FCAT EQUATING AND ITS IMPACT ON STUDENT, SCHOOL, AND DISTRICT DEVELOPMENTAL SCALE SCORES FOR 2003 THROUGH 2010

Introduction

This report is one in a series of special reports produced and commissioned by the Florida Department of Education (FDOE) in response to district concerns that the Florida Comprehensive Assessment Test (FCAT) results for 2010 could contain an error somewhere in the processing. This report will attempt to provide some numerical context for assessing the overall quality of FCAT equating for 2003 to 2010 for Grade 3 through Grade 10. Because of the state-wide perception that there may be particular issues for equating from 2009 to 2010 for the elementary grades, the report will provide additional details for 2009 and 2010 equating for Grades 4 and 5 which directly affects the estimates for student growth between these two grades between 2009 and 2010. In addition, because of districts' concerns about score gains at different levels of achievement, the report will move beyond means across all students and will examine scores and score changes for students with varying FCAT scores.

This report should not be read alone, but should be read with other documentation being released by FDOE. Furthermore, it should be made clear that HumRRO was an active participant in the statistical equating procedures for 2009 and 2010, as it has been since 2000. The work of this report represents a much deeper level of post-hoc analyses that is never possible during the time-compressed period in which operational equating is conducted. HumRRO appreciates the opportunity to review in-depth its prior work and the historical consequences of that work. Hindsight cannot correct history, but it can well prepare us to a better future.

Equating Fundamentals

FCAT for reading and mathematics began in 1998 in one elementary, one middle school, and one high school grade. In 2001, the remaining grades were added such that reading and mathematics is now assessed in all grades from Grade 3 to Grade 10. Subsequently, FCAT incorporated science tests in Grades 5, 8, and 11. While results for reading FCAT in Grades 3 and 10, and in mathematics in Grade 10, have high stakes for students, students' scores for all assessments have high stakes for schools and districts. Along with the high stakes come issues of comparability of test scores. For any grade or subject, any particular student with any particular level of accrued grade-level knowledge and skill should receive the same FCAT score regardless of which year the grade-level test form came from.

Equating is the rather complicated procedure used to ensure that test scores are comparable across years. Equating involves multiple steps and multiple decisions, but basically the idea is to include in each new test form a collection of items that were previously administered. Because of their role in assuring score comparability, the repeated items are referred to as "anchor items." These anchor items have a known relationship to the number scale on which FCAT results are reported, i.e., FCAT's Developmental Scale (DS). When these repeated items are embedded with new items, the repeated items can be used to determine the relationship of the new items to the FCAT DS. This allows students to be assigned DS scores (DSS) that are expected to be equivalent across years.

While simple in concept, the execution is complicated by a variety of test construction principles, with the selection and placement of anchor items being primary issues. The representativeness of these items is crucial, but at the same time it is limited by the pool of items available to act as anchor items and by the selection of non-anchor, scored items. For mathematics and for science, the selection and placement of anchor items on a test form is done for individual items. For reading, however, items are associated with passages, and therefore selection of anchor items must be done by the selection of passages. This adds a constraint for the equating of reading that equating for mathematics does not have to contend with. As a result, equating for reading has historically been more challenging than equating for mathematics.

In addition to the selection and placement of anchor items, creating comparable scale scores across forms involves the use of a set of statistical procedures referred to as Item Response Theory (IRT). These procedures recognize that items differ in difficulty and that it is almost impossible to create multiple test forms for which the sum of correct points (i.e., total raw score) on one form carries equivalent meaning in terms of knowledge and skill to the sum of correct points on other test forms *for all possible total raw scores*. IRT resolves this dilemma by using probabilistic, statistical methods to place items as well as students on the FCAT Scale. With differences in item difficulties captured by this IRT methodology, varying sets of items can be administered to students. Then, appropriate mathematical computations can be used to produce student test scores that take into account variations among forms in the difficulties of the items that they contain. The basic point here is that while these procedures provide a powerful tool for overcoming our lack of full and precise control over the raw score difficulty of new test forms, the procedures are statistical, probabilistic procedures that are subject to some degree of remaining uncertainty.

All of this has been said to alert the reader to the fact that test equating across years, while arguably quite precise, is not perfect but does contain some amount of uncertainty that is extremely difficult to estimate. The factors that create uncertainty act to produce scale scores that are likely be higher or lower than they might be if equating could be perfect. *The issue is not whether equated scores are too high or too low because that is essentially a given. The issue is with the magnitude of the uncertainty and when "too high" or "too low" becomes unacceptably too high or unacceptably too low.*

Florida now has had two episodes in which the equating results have generated skepticism about the precision of FCAT equating procedures. As might be expected from our knowledge about generic staff reactions to unfavorable performance ratings, that skepticism has risen at times of lower than expected assessment scores. It is also true by the nature of the uncertainty of our equating procedures, that if, in any given year, score estimates are a little higher than they should be, the system is designed to self-correct, so that in the following year scores will tend to be back on track. That will create a situation, however, in which if gains in test scores are "too" high the first year because of equating fluctuations, then the gains in test scores may be "too" low the following year. It is this type of "bump in the road" that becomes noticeable. Thus, with the state-of-the-art methods used by FCAT, the high estimate of the first year is only detectable in hindsight.

Student-Level Measurement Uncertainty

It is well known that all ability assessments are only samples of an individual's performance on a relatively small collection of assessment items taken at a particular point in time. Had a student been given other items on a different day, the estimate of his or her knowledge and skill could very well be different. Statistical test theory has recognized this uncertainty and has created various indices to communicate the amount of confidence we can have in student-level assessment scores. Students, teachers, and parents rely on student scores, and rightfully can become concerned when test equating is questioned. In order to understand the magnitude of equating uncertainty, it is important to have some idea about the size of FCAT's student-level measurement uncertainty.

We need to begin with the DS itself. Although the highest score is 3000, that range is for all assessed grades (Grades 3 though 10). No single grade-level assessment covers the full range. For the Grade 4 reading test, for example, the reportable DS scale score range is approximately 300 – 2640 and for Grade 5 reading it is 475-2715. That is roughly 2300 points between the lowest and highest possible scores on these tests. This range sounds impressive; in actuality, the range of possible points has no implication for the precision of the scores.

The measurement certainty of ability tests built with IRT methods is more appropriately expressed as the technical index "*conditional standard error of measurement" (CSEM)*. "Error" is an unfortunate choice of words for this testing statistic. "Error" in measurement terminology does not mean a mistake. It means uncertainty.

Omitting some methodological nuances, CSEM can be interpreted as an estimate of the range of scores for which we have some degree of certainty. For example, if a CSEM is 100, then we believe that we are within 100 points of what a score should be about two-thirds of the time. Under the methodology of IRT, tests tend to be built to be most accurate in the middle of the distribution of student abilities. FCAT divides the achievement into five levels which requires four cutpoints on the DS. (A cutpoint is the location on the DS where the score associated with achievement changes from Level 1 to Level 2, Level 2 to Level 3, Level 3 to Level 4, and Level 4 to Level 5.) Keeping in mind that the DSS contains approximately 2300 score point for most grades, the CSEMs at the cutpoints, going from low to high, are 82, 70, 70, and 105 DSS points for Grade 4 reading in 2009 and 95, 79, 84, and 123 DSS points for Grade 5 reading in 2010. CSEM values vary across years, but these examples will serve the purpose of creating context for reviewing equating. Complete CSEM tables at cutpoints for all grades and subjects were produced by FDOE and are reproduced in Appendix A. For an easy-to-remember rule of thumb, reading DSS can be considered certain to about plus or minus 80 scale score points. For mathematics, certainty is better at about plus or minus 40 scale score points.

The History of Scores from 2003 through 2010

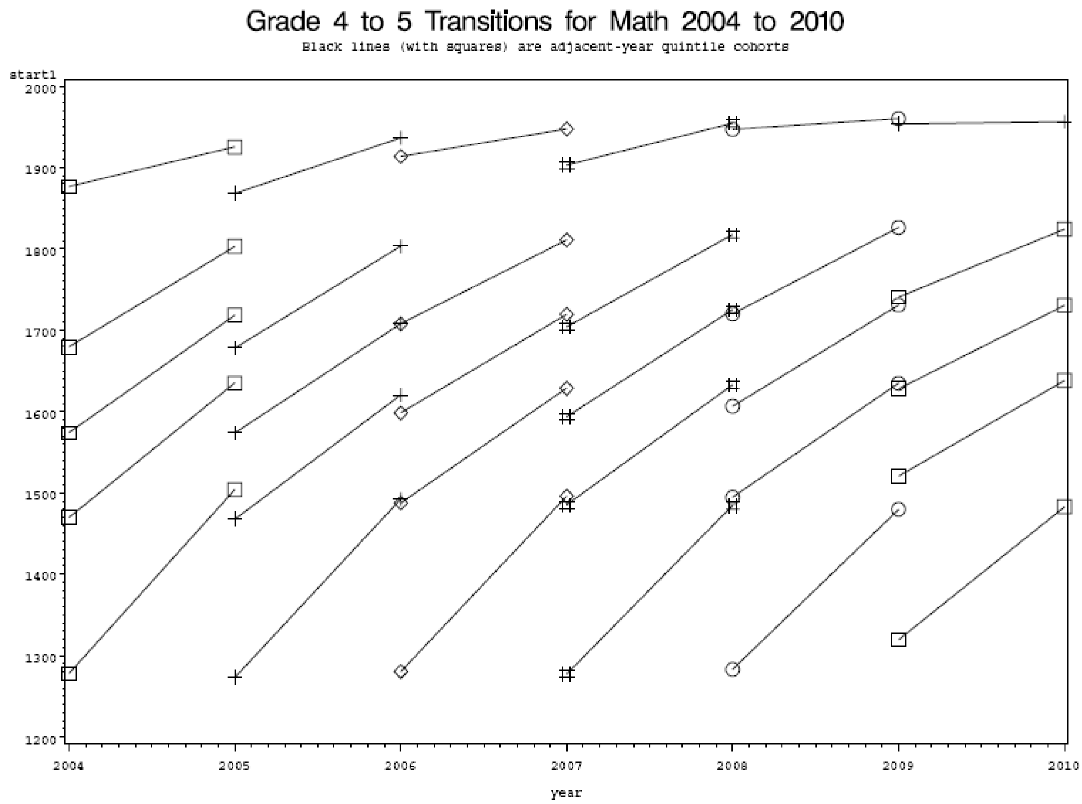
Given that we recognize the equating is not perfect, we prepared graphs showing grade-to-grade and year-to-year scores for cohorts of students. We were hoping for a high degree of regularity in the graphs, suggesting only small random fluctuations in equating.

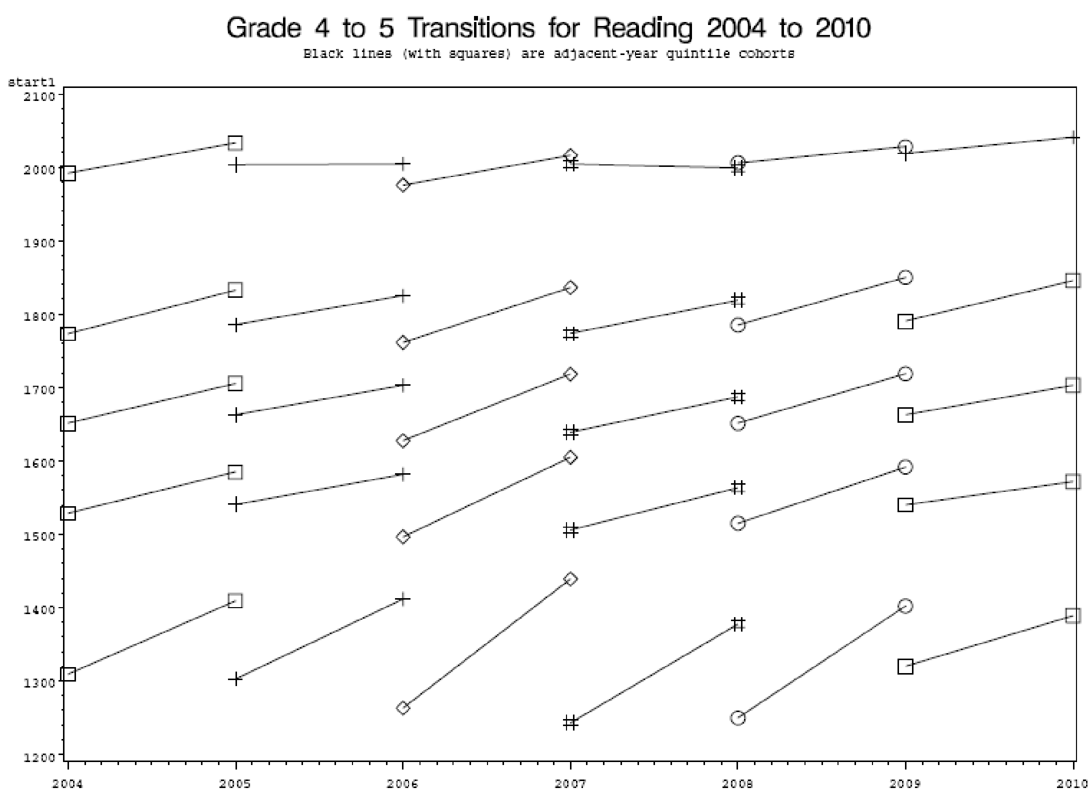
Grade Transition by Academic Year Cohorts

Using 2010 student data files which contain students' current and prior year DSS, we retrieved scores for students who transitioned across adjacent grades from each of the adjacent pairs of years from 2003 to 2010. Graphs for all grade pairs, for reading and mathematics, are presented in Appendix B. For illustration and explanation, the scores for Grade 4 to Grade 5 transitions for reading and mathematics are presented below.

Each figure contains what amounts to six columns of line segments. Simply stated, each column represents a pair of years; the first column begins with Grade 4 students in 2004 and ends with the same students in Grade 5 in 2005. The second column begins with Grade 4 in 2005 and ends with the same students in Grade 5 in 2006, and so on. Each line segment within a particular column connects the Grade 4 DSS mean to the Grade 5 DSS mean for the same students. The lines in the first column connect mean FCAT scores in 2004 with mean FCAT scores in 2005. The lines in the second column connect means for the next cohort of students who have Grade 4 FCAT scores in 2005 and Grade 5 FCAT scores in 2006, and so on. For each pair of years, students were divided into five equally-sized groups, called quintiles, based on their Grade 4 score (Levels 1-5). As a result, the first quintile contains the 20% of the students with the lowest FCAT scores, and so on. The reader will note that each column shows five line segments; these five line segments represent the quintiles for each pair of years. The line segment at the bottom of the column represents the lowest performing quintile of students; the line segment at the top represents the highest performing quintile for that particular pair of years.

These graphs contain a wealth of information. First, knowledge gains between Grade 4 and Grade 5, as represented by the vertical difference on the DSS scale between the beginning and ending of a line segment, is smaller than the score ranges within either grades. Furthermore, for mathematics, the lowest quintile of students tends to finish Grade 5 with a DSS that is roughly where the second lowest quintile ended Grade 4. For reading, the lower quintile of students tends to finish Grade 5 below the point at which the next higher quintile finishes Grade 4 In other words, within a grade, differences between quintiles appear about the same as differences between grades within a quintile. From this perspective it is difficult to determine what defines "on-grade" performance.


Grade 4 to 5 Transitions for Math 2004 to 2010
Black lines (with squares) are adjacent-year quintile cohorts
start1
2000
1900
1800
1700
1600
1500
1400
1300
1200
2004
2005
2006
2007
2008
2009
2010
year


Grade 4 to 5 Transitions for Reading 2004 to 2010
Black lines (with squares) are adjacent-year quintile cohorts
start1
2100
2000
1900
1800
1700
1600
1500
1400
1300
1200
2004
2005
2006
2007
2008
2009
2010
year

Second, differences in DSS means between grades are greater for the lower quintiles than for the upper quintiles. In other words, students in the lowest quintile have a greater mean difference between Grades 4 and 5 than students in the highest quintile. In fact, the highest level students appear to be hardly growing at all.

Both of the above observations need to be acknowledged, although neither is related to the issue of equating. In terms of equating, we would prefer the above patterns to be smooth and regular across all year-pairs. Comparing mathematics to reading, we see what we would expect to see. The "flow" of the line segments is more regular for mathematics than for reading. That is, for mathematics, for each quintile, the line segments across years are reasonably parallel. That is not the case for reading, particularly for the lowest quintile. For that lowest quintile, the starting and the ending points for each line segment vary more that any other quintile. Given districts' concerns about 2009 and 2010 growth, the 2009 to 2010 line segment for the lowest quintile does appear atypical compared to previous years. The magnitude and impact of that atypical line segment will be a focus of attention for the remainder of this report.

Presenting the data in this way is somewhat of a breakthrough for the industry. Notice that the line segments for the middle quintile for both reading and mathematics are reasonably stable, particularly since 2007. The industry tends to monitor trends only for the center of the distribution of students since this is the group of students the test is best designed to measure as it encompasses the majority of students. Modifications will be made for future FCAT quality assurance monitoring to track equating projections across the range of scores so that students performing at the highest and lowest level are also monitored.

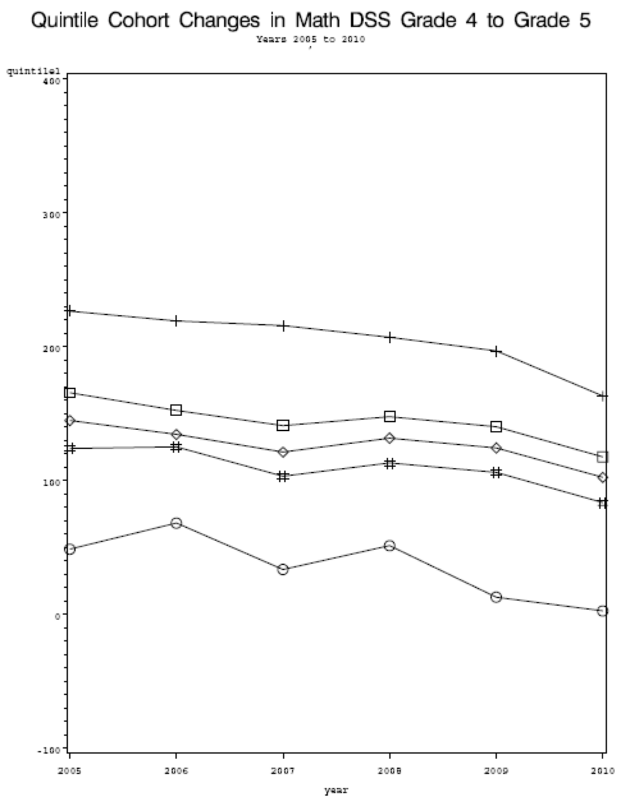
Again, similar graphs for the remaining reading and mathematics grades appear in Appendix B.

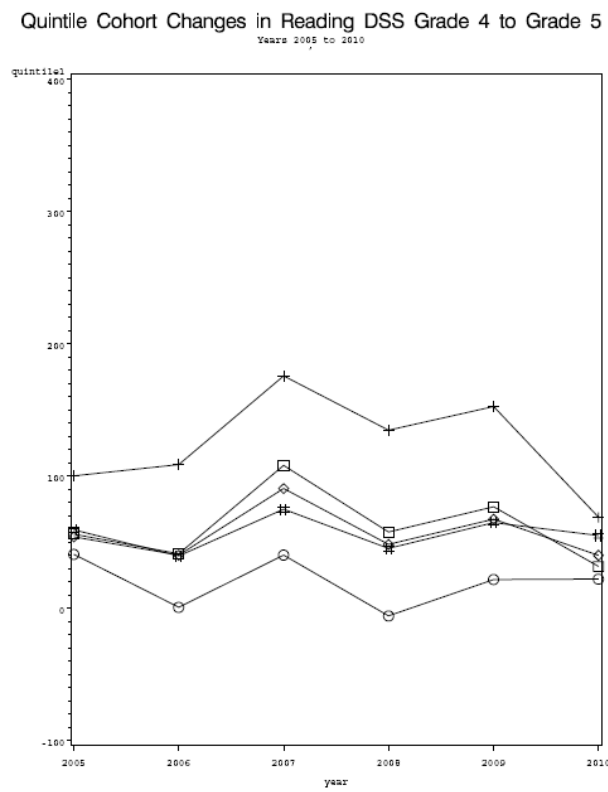
Score Gain Histories 2003 to 2010

We also used the files generated for the above graphs to directly track differences in individual student growth across adjacent years. This information is presented in the following graphs. For each quintile group, the average DSS change from one grade to the next was calculated and graphed so that the mean to the farthest left is the mean DSS change of students from Grade 4 to Grade 5 in 2005, the mean DSS change of students from Grade 4 to Grade 5 in 2006, etc. Again, for mathematics there are no cross-year abrupt changes like there appear to be for reading. Even for mathematics, however, the estimation of DSS change is less predictable for the lowest quintile than for the other quintiles.

Notice, also, that the gains, i.e., changes in DSS, between years for mathematics are declining. Reading is harder to interpret in this regard. On the other hand, the reading graph suggests that from 2007 to 2009, the general trend for reading gains is downward for all except the highest quintile of students. If we accept this premise and focus on the gains for the lowest quintile, we can project from the general downward trend that the gain for 2010 could have been 100 DSS points rather than the 80 points that are observed. That clearly is a very rough estimate, but it suggests the potential underestimate of 20 DSS points (80 versus 100) of the average gain for students in the lowest quintile. Twenty points is about 1/5th of the standard error of measurement for individual students in the DSS range of the lowest quintile. In the context of the DS, 20 is a small number. If it does represent an abnormal equating fluctuation, it should have no practical impact on interpreting scores for individual students. Scores in the DSS range of the lowest quintile cannot be considered accurate to plus or minus 20 points.

Similar graphs for the remaining reading and mathematics grades appear in Appendix C.





Relationships between Scores across Grades

Another way to examine the stability of equating is to examine the relationships between scores across adjacent grades by student cohort. To do this, we use the Stocking/Lord procedure, a statistical procedure used to place items from two different forms onto the same scale so that scores from the two forms are comparable. The end result is a linear translation which uses the slope and intercept of a linear equation to place the forms on the same scale.

Student-Level Analysis

Technically, equating fluctuations can occur in both the slope and intercept of the linear equation that Stocking/Lord produce to equate test forms across years. The slope and intercept will play out in different ways on the relationship between DSS across two adjacent years. Fluctuations in the equating intercept will cause a regression line used to predict DSS in grade X from DSS in grade (X-1) to move up and down (a main effect for year of the equating) and fluctuations in the equating slope will cause the regression line to be more or less steep (as the variance of grade X is expanded or compressed by equating).[1]

A file was created across years with 728,597 students containing their Grade 4 reading DSS, their Grade 5 reading DSS, and the year in which they were in Grade 5.[2] The technical results are that the multiple R-squared predicting of Grade 5 DSS from Grade 4 DSS alone is 0.5764. Year of equating was then added as a categorical main effect to look at the stability of equating intercepts over time. Year of equating increased the multiple R-squared to 0.5799. The interaction of year of equating and Grade 4 reading DSS was then added to look at the stability of equating slope coefficients over equating years which increased the multiple R-squared to 0.5820.

Practically speaking, one can interpret the above results in the following way. If a student has a score of *X* in Year 1, he or she will be predicted to have a score of *Y* the next year with a reasonable degree of certainty. We can make our prediction based on the relationship between Grade 4 and Grade 5 scores across all students and all years, or we can make our prediction based on relationships across all students, but tailored to each pair of adjacent years. If unintended equating fluctuations are minimal and if there are no outside influences on students or schools (e.g., no hurricanes, no recession) and schools are simply maintaining the status quo (not improving the amount that students improve each year), then one overall prediction equation should work as well as prediction equations tailored for each year. That is, if a student has a score of *X* in Grade 4, and a "one-size-fits-all" equation predicts *Y*, while a tailored, year-specific equation predicts *Z*, when all of the conditions above are met, then *Y* and *Z* should be close to each other. The multiple R-squared results tell us the difference between *Y* and *Z* is likely to be within 20 points on the DS.[3] The DS per se is not precise enough to pick up a 20-point potential difference in any one student's score. On the aggregate, however, it may be meaningful.

The difficult part is making the attribution that the difference across years is equating uncertainty or real structural conditions affecting schools and/or students. We conducted the above analysis in an alternative manner to see if schools in Florida were, on average, improving their capacity to teach students more each successive year with DSS increases being at a constant rate across years. If that were happening, then year-to-year differences in predictions would show a pattern in which, for each successive year, students would need to have some constant DSS points added to their predicted score. That was not the case. The year-to-year prediction

---

[1] This assertion depends on correlations between DSS in grade X and DSS in grade X+1 to be essentially constant across years. Since we are examining the Grade 4 to Grade 5 transitions from 2004 to 2010, we have six observation periods. For those six time periods, within-cohort correlations were .77 for three of the time periods, .76 for two time periods, and .74 for the remaining time period.

[2] Again, the data was created from the student data files for 2010. For students in Grade 10 this year, they moved through the Grade 4 to Grade 5 transition in between 2004 and 2005. Given the time span, not all students whothat successfully transitioned between Grade 4 and Grade 5 between 2004 and 2005 are represented in the data set created from students in Grade 10 in 2010.

[3] The standard deviation for Grade 5 reading across all years from 2005 to 2010 is 285. The increase in R-squared that results by adding year in the equation is about one-half a percent of the total variance in students Grade 5 scores, or about DSS 20 points.

differences are not regular. The source of the 20-point difference in cross-year predicting remains any unknown combination of factors, including desirable factors such as real changes in schools' capacity to teach students more in successive years, unfortunate nuisance factors like instructional days lost to hurricanes, and potentially some unintended test equating issues. From this perspective, 20 points may be viewed as an upper bound on equating uncertainty; it could be much less than that.

We also attempted to predict student gains between Grades 4 and 5 using initial Grade 4 DSS and cohort year. The multiple R-squared started at 0.0644, went to 0.072 with the addition of year, and stopped at 0.077 with the addition of the year-by-starting point interaction. Practically speaking, we cannot predict individual students' gains from their starting points.

Results for the remaining student-level grade transitions for reading and for mathematics appear in Appendix D. They appear at least as robust as the Grade 4 to Grade 5 example presented above.

School-Level Analysis

For the first time in the report, we introduce our examination of school-level results. We conducted the same analyses as above for schools; that is, we examined the extent to which the average of students' DSS in one grade could predict the average of students' DSS the following year in the next higher grade. In order to remove extraneous variations caused by potentially shifting student populations, only students who made the transition between the two grades were used to calculate grade/year average scores for schools.

Across the years 2004 to 2010, schools' Grade 5 average DSS can be predicted by Grade 4 average DSS for the prior year with an accuracy expressed by the R-squared of 0.8702. Adding the year in which students completed Grade 5 increased the R-squared to 0.9011, and in the interaction of year and grade increased the R-squared to 0.9041. The difference between the R-squares with and without considering year is small, but does translate, once again, to about 20 DSS points. That is, if we (a) use the general trend across all schools and years to predict a school's Grade 5 average for its students in one particular year based on the average for those same students on Grade 4 in the prior year, and then (b) use the general trend across schools but tailored to the particular year to predict a school's Grade 5 average for its students in one particular year based on the average for those same students on Grade 4 in the prior year, and then (c) compare the accuracy of the first prediction to the accuracy of the second prediction, we will find that we are more accurate by about 20 points with the second method. Once again, some combination of factors, including desirable factors such as real changes in schools' capacity to teach students more in successive years, unfortunate nuisance factors like instructional days lost to hurricanes, and potentially some unintended test equating issues, combine to create this 20-point variation in prediction accuracy.

A difference between school- and student-level analyses is that, while we can assume that the school-level average is a more reliable measure than the student-level scores, we do not yet have available a standard error estimate for school scores, although one can be produced by a methodology called Generalizability Theory. Also, the standard deviation for school averages is about a third of the standard deviation for student scores since the number of schools is smaller than the number of students, so the 20-point different in year-to-year predictions is more noticeable at the school level of analysis than at the student level of analysis. As with the student-level analysis, we can interpret the 20 points as an upper bound for the possibility of equating uncertainty and non-test factors (e.g., changes in school capacity may also be represented in those 20 points).

For schools, the prediction of average gains in students' scores was also examined. Schools' Grade 4 scores alone were not related to average gains for students within schools (the multiple R-squared was 0.0002). On the other hand, when year was added to the prediction, the R-squared increased to 0.2380 for the main effect and 0.2614 with the inclusion of the interaction. This is a larger increase in predictability by year than the previous analyses. Interestingly, because the variance in school gains is relatively small (33.6), the impact of year is still about 20 points.

Results for the remaining grades and subjects appear in Appendix E. As with the student-level results, they appear at least as robust as the Grade 4 to Grade 5 example presented above. The school results continue to support our rule of thumb that 20 points on the DSS scale is a reasonable upper bound estimate for potential equating fluctuations. As a reminder, we use the term upper bound to mean that if all of the 20-point fluctuation across years is attributed to equating processes per se, then there is no room for attributions about schools' changing their capacity to improve instruction from year to year.

District-Level Analysis

The analyses were repeated with districts' averages of DSS for their students. Across the years 2004 to 2010, districts' Grade 5 average DSS can be predicted by Grade 4 average DSS for the prior year with an accuracy expressed by the R-squared of .7803. Adding the year in which students completed Grade 5 increased the R-squared to .8858, and in the interaction of year and grade increased the R-squared to .8898. The difference between the R-squares with and without year translates to about 23 DSS points based on a Grade 5 district-level standard deviation of 70 across all years.

For district gains, the prediction of average gains in students' scores was also examined. Districts' Grade 4 scores alone were not related to average gains for students within schools (the multiple R-squared was .0019). On the other hand, when year was added to the prediction, the R-squared increased to .4815 for the main effect and .4914 with the inclusion of the interaction. This is a larger increase in predictability by year than the school analysis. On the other hand, variance in district gains is only 22.25 DSS points, so the impact of year is about 16 DSS points.

Results for the remaining grades and subjects at the district level of analysis appear in Appendix F. Once again the remaining transitions appear at least as robust as the Grade 4 to Grade 5 example presented above.

Conclusion from Regression Analyses

Our conclusion from these analyses is that the 20-point (plus or minus) DSS variation in prediction accuracy can be interpreted as an upper bound of unintended equating issues. Some of that 20-point DSS variation may be due to real structural changes, not in the test, but in the students and schools. And, to repeat, given the range of the scale and the measurement uncertainty within a given year, the potential is miniscule for a 20 DSS point equating uncertainty at the student level of analyses to have a practical impact.

Revised Equating

In one of the companion FDOE reports, Dr. Hill has pointed out that changing the test positions of passages used for equating. For equating Grade 4 reading in 2009, one anchor passage was placed in the test form one position away from its position in its prior administration. In addition, in 2010 there were two anchor passages in Grade 4 2010 that had been shifted on position away from its prior administration. We want to acknowledge Dr. Hill's concern.

Because equating for 2010 depends on the results for 2009, we first re-equated Grade 4 for 2009 and recomputed students' DSS scores. We then re-equated 2010 for Grade 4 and rescored those students. This reanalysis was limited to Grade 4 simply due to time constraints.

The following table shows DSS scores with the original equating and the alternative equating. On average, re-equating lowered scores for both 2009 and 2010. With both original and revisited equating, there was a 4-point mean drop in Grade 4 reading DSS. However, the change in equating was not uniform. With either computation, students who scored one standard deviation below the mean lost more points: 30 DSS points for the original equating and 14 DSS points with the revisited equating. Students who scored one standard deviation above the mean, however, gained points: 22 DSS points with the original computation and 6 DSS points with the revisited computation. As a result, the revisited equating added points to the 2009-to-2010 change in scores for students at the lower end of the distribution, and removed points for students at the high end of the distribution. Once again, we have arrived at a number for potential equating fluctuations roughly around 20 DSS points, but in opposite directions dependent on score level.

Grade 4 Reading Equating			
2009 DSS			
	-1 sd	Mean	+1 sd
Original	1298	1605	1913
Revisited	1281	1594	1907

2010 DSS			
	-1 sd	Mean	+1 sd
Original	1268	1601	1935
Revisited	1267	1590	1913

Difference Orginal and Revisited			
	-1 sd	Mean	+1 sd
2009	-30	-4	22
2010	-14	-4	6

Difference in equating results			
	-1 sd	Mean	+1 sd
	17	0	-16

Impact of Field-Test Items

An alternative hypothesis for perceived score declines suggested by the districts related to the field-test items placed on the 2010 assessments. The concern was that field-test items could disrupt students, leading them to score lower on the scored items than they otherwise would. For 2010, we have the opportunity to directly test that question.

For Grades 3, 4, 5 and 9, the anchor forms and field-test forms were spiraled together within schools for all schools in Florida. For these grades, we essentially have a random assignment of students to forms. Therefore, students were divided into two groups -- those who received an anchor form versus those who received a field-test form -- and differences in their FCAT DSS score were examined. The “statistical P” column in the table below shows the results of statistical significance tests for differences in DSS for the two groups. Effect sizes are also shown. Effect sizes compare the difference in scores to the standard deviations of the groups. The large number of students in Florida leads to significant differences even when the differences themselves are very small, and thus not practically significant.

All grade/subject are displayed, but again, only the grades in bold received state-wide randomly spiraled test forms this year. Effect sizes are quite small; there appears to be no cause for concern about the field-test items disadvantaging student performance on the operational, scored items.

	Reading		Mathematics	
Grade	Statistical P*	Effect Size**	Statistical P	Effect Size
3	**.5933**	**+0.001**	**.0975**	**-0.014**
4	**.1423**	**+0.000**	**.0105**	**+0.036**
5	**.0001**	**+0.044**	**.0272**	**-0.013**
6	.0073	-0.005	.0001	-0.022
7	.0108	-0.026	.0001	-0.044
8	.0001	-0.040	.0001	-0.036
9	**.0001**	**+0.069**	**.0001**	**+0.072**
10	.0309	-0.001	.5999	-0.003

*SAS Proc GLM testing developmental scores for students administered field test form versus anchor form, controlling for curriculum (standard versus not standard curriculum).
**Mean for field test form students minus mean for anchor form students.

Note that the above table also supplies a test of DSS mean differences between the students who responded to anchor items and were therefore involved in the equating process versus those students who did not produce data for the anchor items. The two groups of students are not practically different. To provide a point of reference, a .04 effect size is approximately 11 DSS points.

Conclusions

First, HumRRO wishes to express it's appreciation for being asked to participate in addressing the concerns of the Florida school districts regarding FCAT scores for 2010. Although we believe that Florida has a strong system for quality control, test equating is complicated and must be conducted under tight time pressures. Therefore, when we hear criticisms we always have what we consider to be a healthy fear that some part of the process has gone awry.

Second, given the time that we have spent in the last few weeks reviewing FCAT score results and conducting new analyses to review data in historical context, we are confident that no mistakes have been made. The data are simply too regular and too predictable. It has even become predictable that there will be fluctuations in lower level gains for reading.

Third, equating by its nature is not exact. Our data analyses, surprisingly, kept returning what we have called "upper bound" estimates of the uncertainty of equating at about 20 Development Scale Score points for Grades 4 and 5 reading. On the DS, this is a small number and a number that is noticeably smaller than basic measurement uncertainty. It would be impossible to qualitatively describe knowledge and skill differences between two students with Developmental Scale Scores only 20 points apart.

Fourth, average Development Scale Scores for schools should, in theory, take advantage of the fact that measurement uncertainty for students is typically regarded as random – meaning that some students will receive scores that somewhat underestimate their true ability and other students will receive scores that somewhat overestimate their true ability. School- level averages should be closer to the truth. Further research should be conducted, however, to estimate the amount of uncertainty for school means just like we routinely do for student- level measures. That would help place our estimations of school- level score fluctuations in better context. On the other hand, our view of the data provides no cause for concern regarding data processing and equating.

Fifth, because there is a tendency to over interpret small differences, we want to continue repeating the standard psychometric mantra that test scores, whether they are scores for students, scores for schools, or scores for districts, are not perfect. One the other hand, we understand the intense scrutiny that school and district personnel across the nation have come to face under the accountability systems that have emerged during this decade. School reform and improved instruction are difficult. Sometimes the progress is simply too hard to detect with current state-wide assessments. Trends over time should be given more interpretive emphasis than year-to-year changes.

Finally, this has been a learning experience and, we hope, a teaching experience. We are confident that the emerging FCAT-II will be improved by what we now know.

Appendix A

Student Level Measurement Precision:
Statistic Probabilities Expressed as *Conditional Standard Errors of Measure*

Table A-1. Reading—*Conditional Standard Error of Measurement* (Range of Certainty) at Achievement Level Cuts					
3		1046	1198	1489	1866
	2007	97	85	85	158
	2008	91	85	97	152
	2009	85	79	91	152
	2010	97	85	85	146
		1315	1456	1690	1965
4	2007	82	70	76	105
	2008	82	70	76	105
	2009	82	70	70	105
	2010	88	76	82	123
		1342	1510	1762	2059
5	2007	106	78	67	106
	2008	90	78	84	118
	2009	95	78	78	118
	2010	95	79	84	123
		1450	1622	1860	2126
6	2007	89	67	67	122
	2008	89	72	83	122
	2009	89	78	78	111
	2010	94	78	78	111
		1542	1715	1945	2181
7	2007	79	68	79	126
	2008	89	68	73	110
	2009	79	63	68	110
	2010	89	79	79	121
		1696	1882	2073	2282
8	2007	62	48	62	100
	2008	62	52	62	100
	2009	67	52	62	95
	2010	67	62	71	105
		1772	1972	2146	2298
9	2007	76	65	71	87
	2008	76	65	65	87
	2009	65	65	81	109
	2010	76	65	76	103
		1852	2068	2219	2311
10	2007	87	87	92	103
	2008	97	92	92	97
	2009	76	81	92	103
	2010	87	87	92	103

Table A-2. Mathematics— *Conditional Standard Error of Measurement* (Range of Certainty) at Achievement Level Cuts

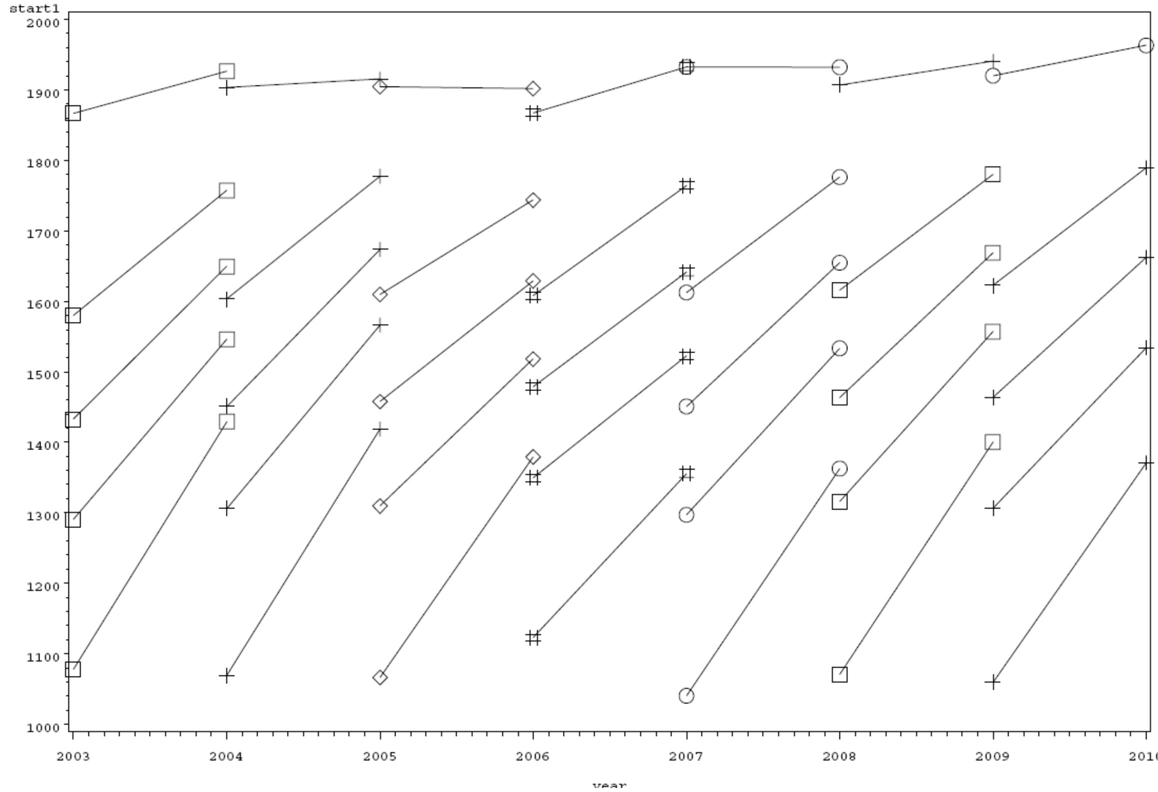
3		1079	1269	1509	1750
	2007	106	79	69	97
	2008	97	69	69	111
	2009	97	74	69	102
	2010	97	79	69	102
		1277	1444	1658	1863
4	2007	101	66	61	83
	2008	92	61	61	96
	2009	83	66	66	92
	2010	79	61	66	87
		1452	1632	1769	1957
5	2007	66	47	42	52
	2008	61	47	42	57
	2009	61	47	42	61
	2010	61	47	42	57
		1554	1692	1860	2019
6	2007	77	56	47	65
	2008	65	56	52	69
	2009	65	56	56	73
	2010	73	56	52	69
		1661	1786	1939	2080
7	2007	65	48	40	52
	2008	61	44	40	52
	2009	61	44	40	56
	2010	61	44	40	56
		1733	1851	1998	2092
8	2007	51	36	32	36
	2008	43	36	32	36
	2009	43	36	28	32
	2010	43	36	28	32
		1782	1901	2023	2142
9	2007	54	34	31	37
	2008	48	34	31	41
	2009	48	34	34	44
	2010	48	34	34	44
		1832	1947	2050	2193
10	2007	53	33	25	29
	2008	45	33	29	29
	2009	45	33	25	33
	2010	41	29	25	33

Appendix B

Grade Transition by Academic Year Cohort

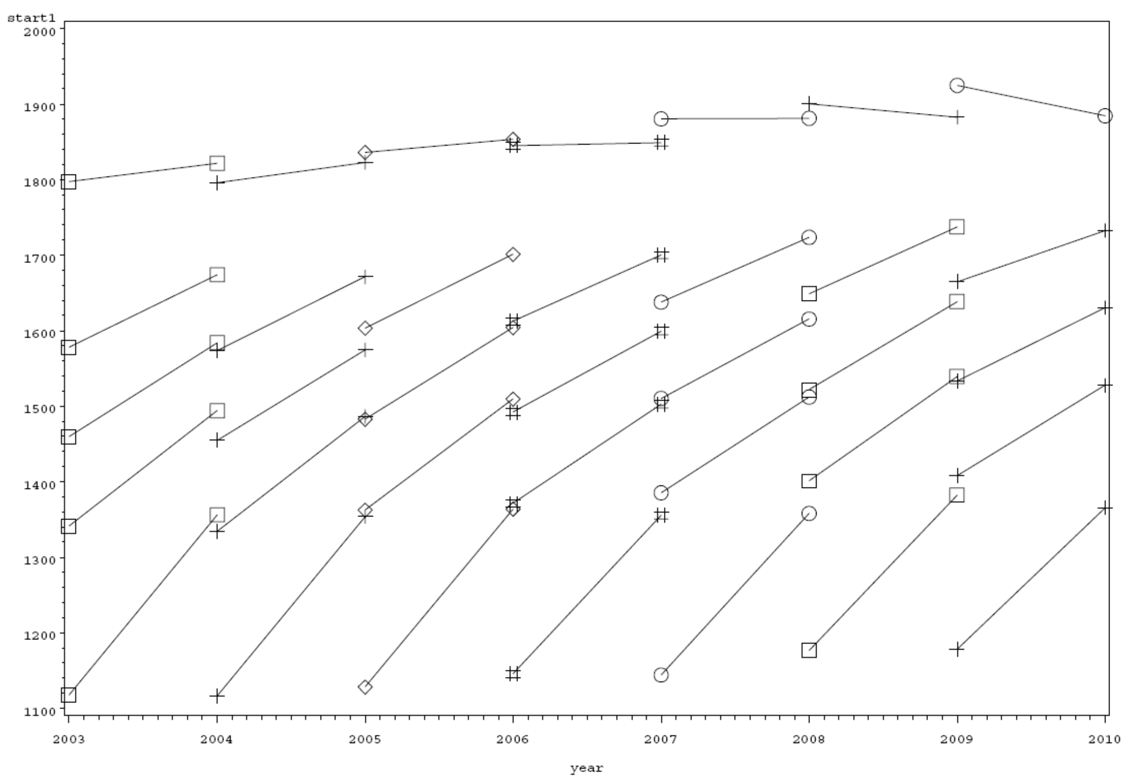
Grade 3 to 4 Transitions for Reading 2004 to 2010

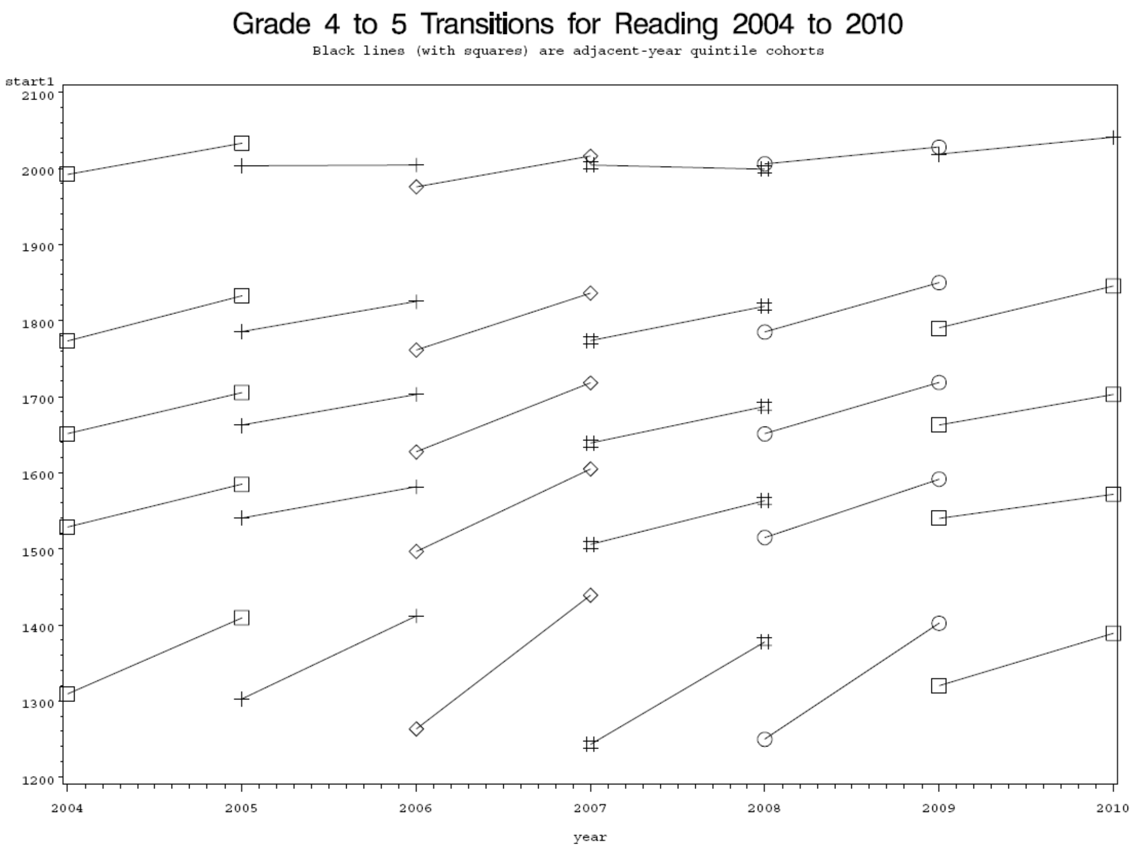
Black lines (with squares) are adjacent-year quintile cohorts



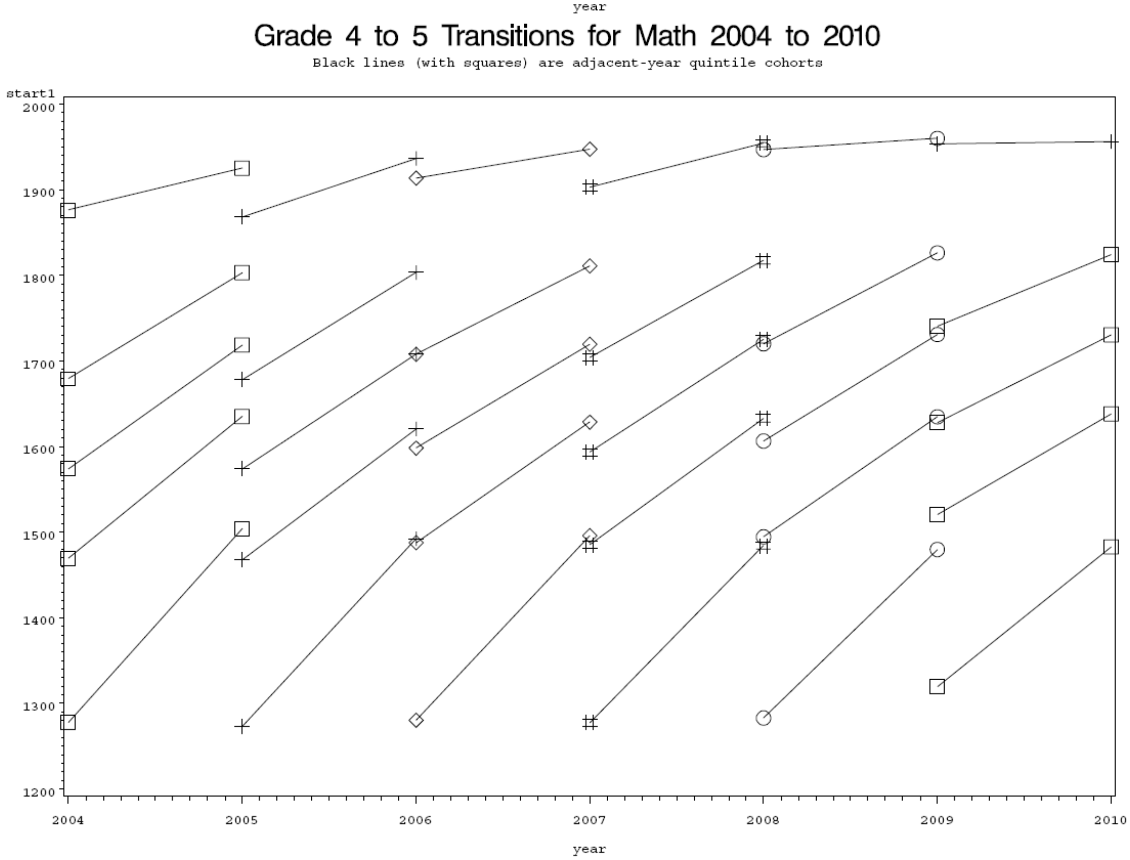
Grade 3 to 4 Transitions for Math 2004 to 2010

Black lines (with squares) are adjacent-year quintile cohorts





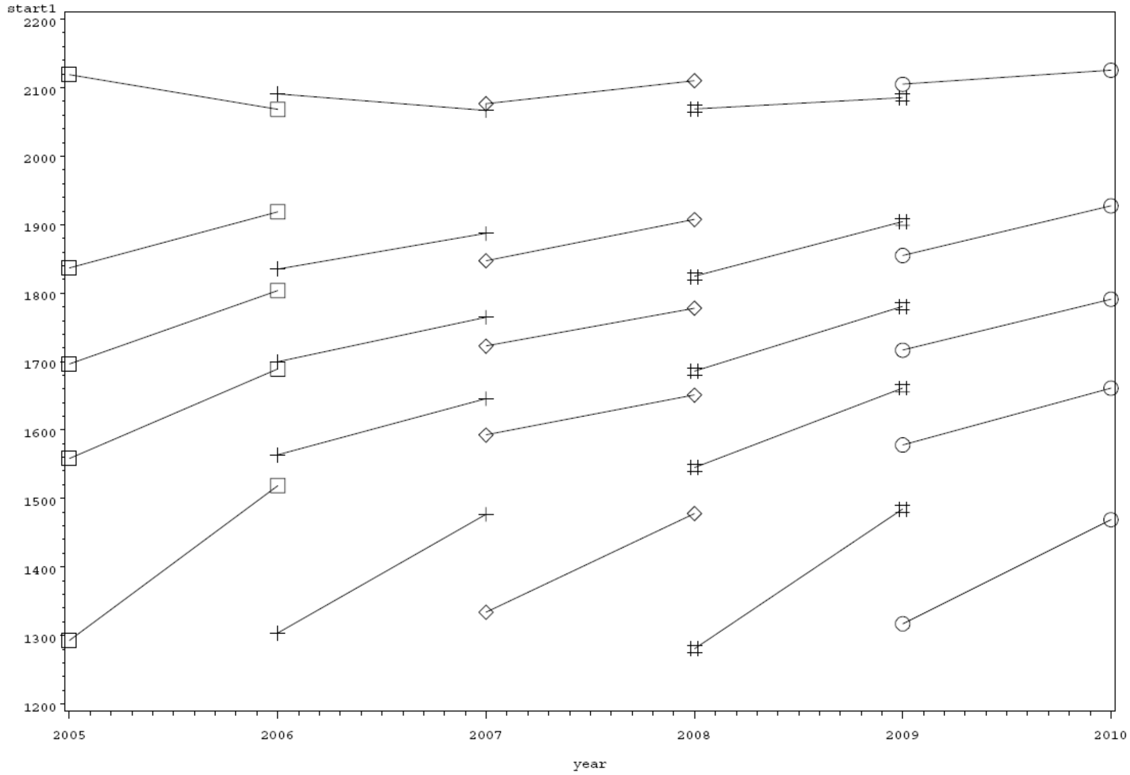
Grade 4 to 5 Transitions for Reading 2004 to 2010
Black lines (with squares) are adjacent-year quintile cohorts
start1
2100
2000
1900
1800
1700
1600
1500
1400
1300
1200
2004
2005
2006
2007
2008
2009
2010
year



Grade 4 to 5 Transitions for Math 2004 to 2010
Black lines (with squares) are adjacent-year quintile cohorts
start1
2000
1900
1800
1700
1600
1500
1400
1300
1200
2004
2005
2006
2007
2008
2009
2010
year

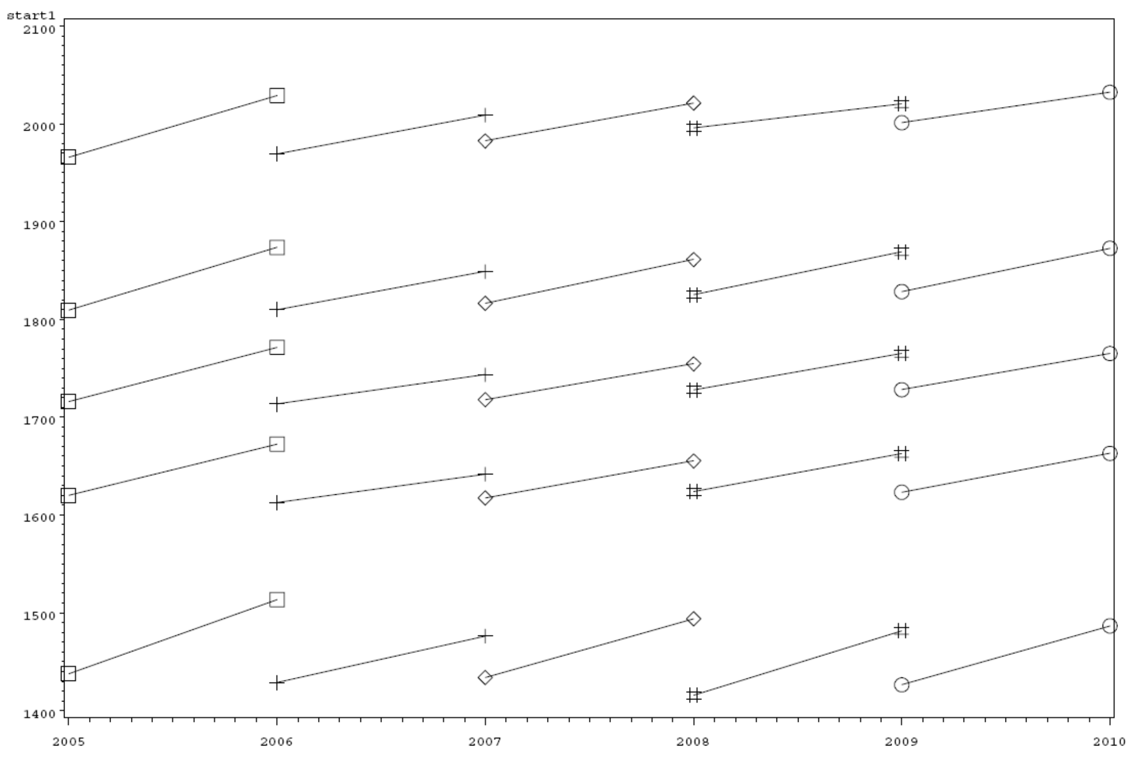
Grade 5 to 6 Transitions for Reading 2004 to 2010

Black lines (with squares) are adjacent-year quintile cohorts



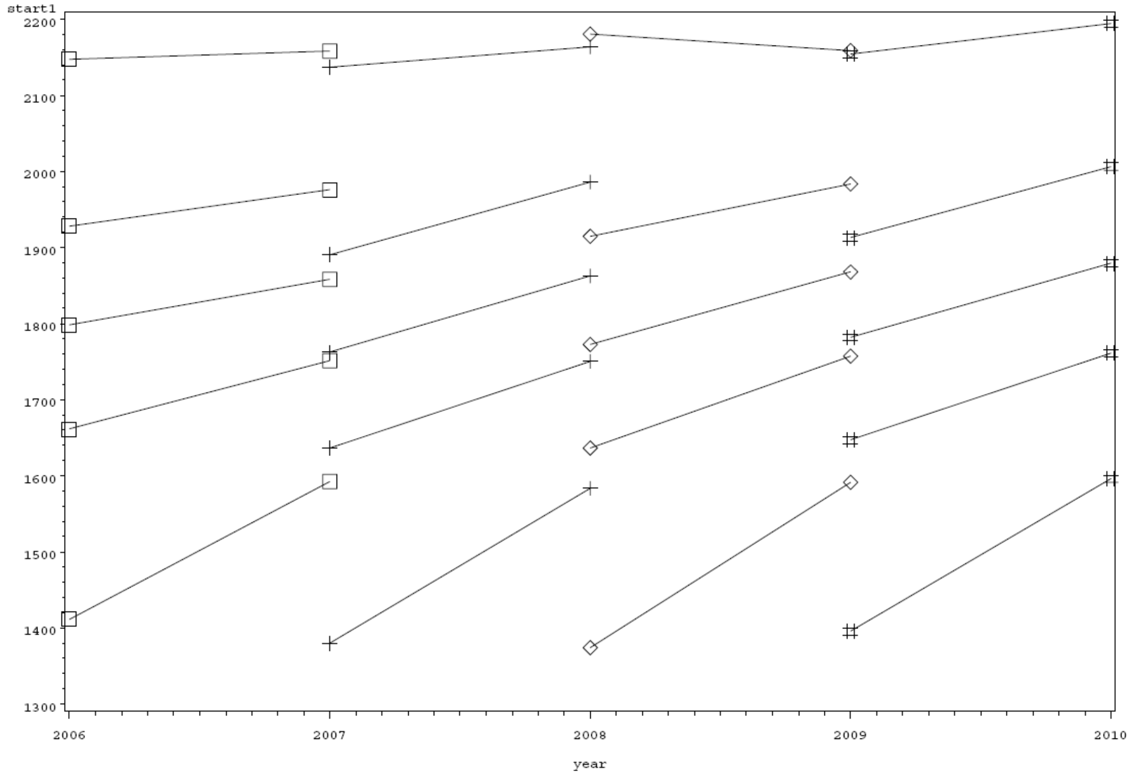
Grade 5 to 6 Transitions for Math 2004 to 2010

Black lines (with squares) are adjacent-year quintile cohorts



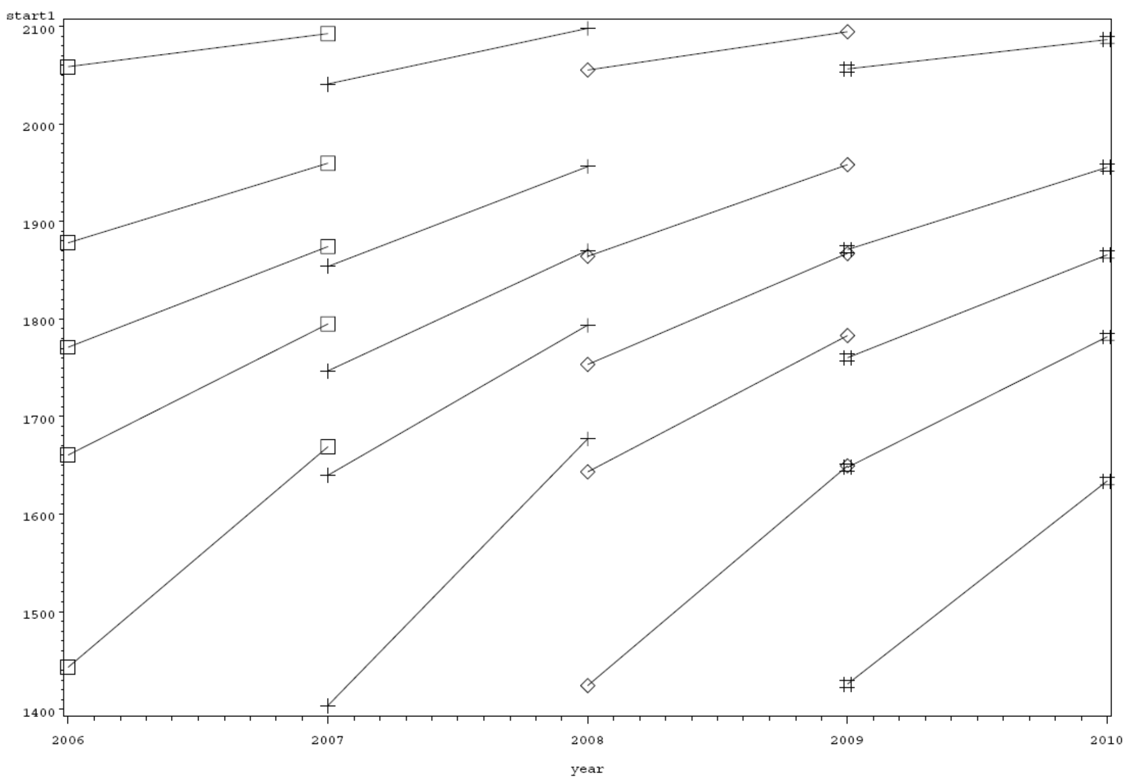
Grade 6 to 7 Transitions for Reading 2004 to 2010

Black lines (with squares) are adjacent-year quintile cohorts



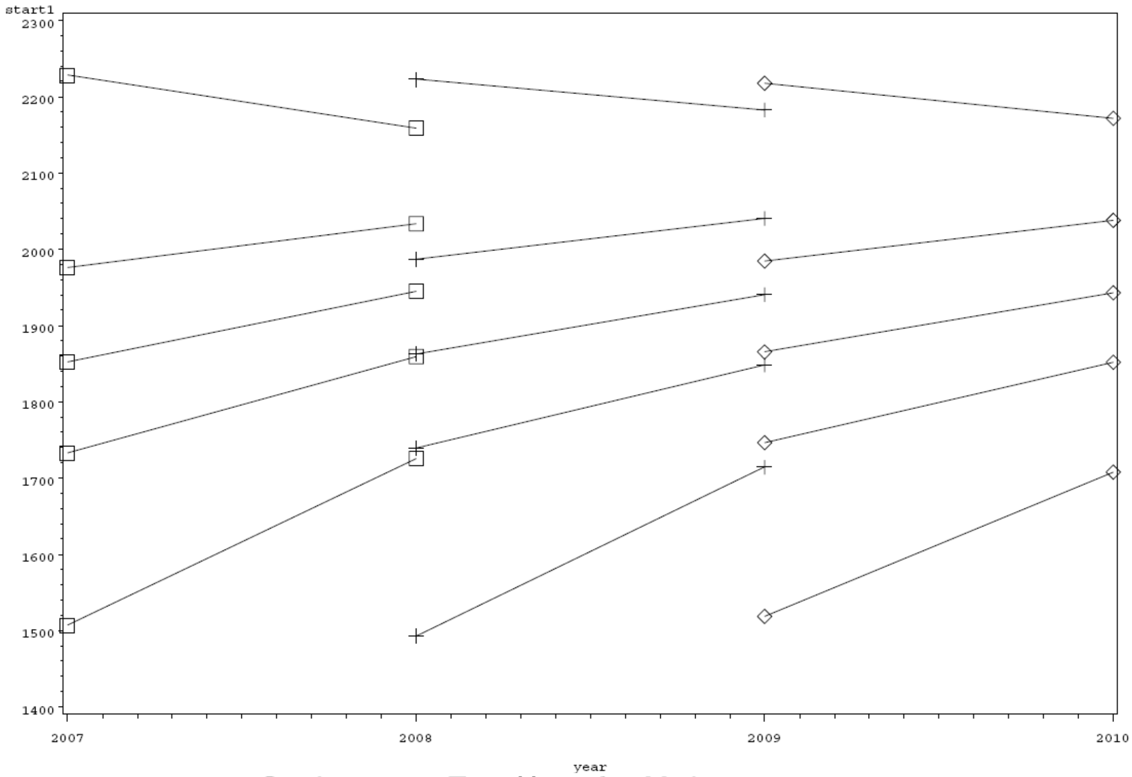
Grade 6 to 7 Transitions for Math 2004 to 2010

Black lines (with squares) are adjacent-year quintile cohorts



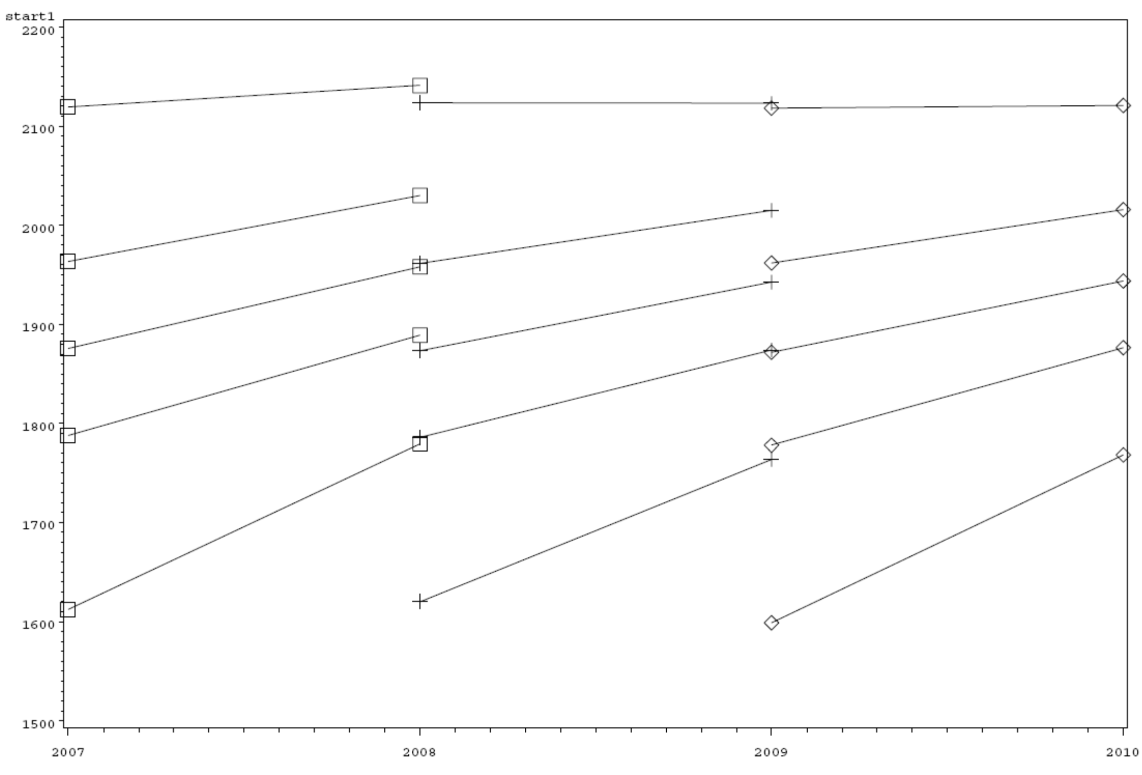
Grade 7 to 8 Transitions for Reading 2004 to 2010

Black lines (with squares) are adjacent-year quintile cohorts



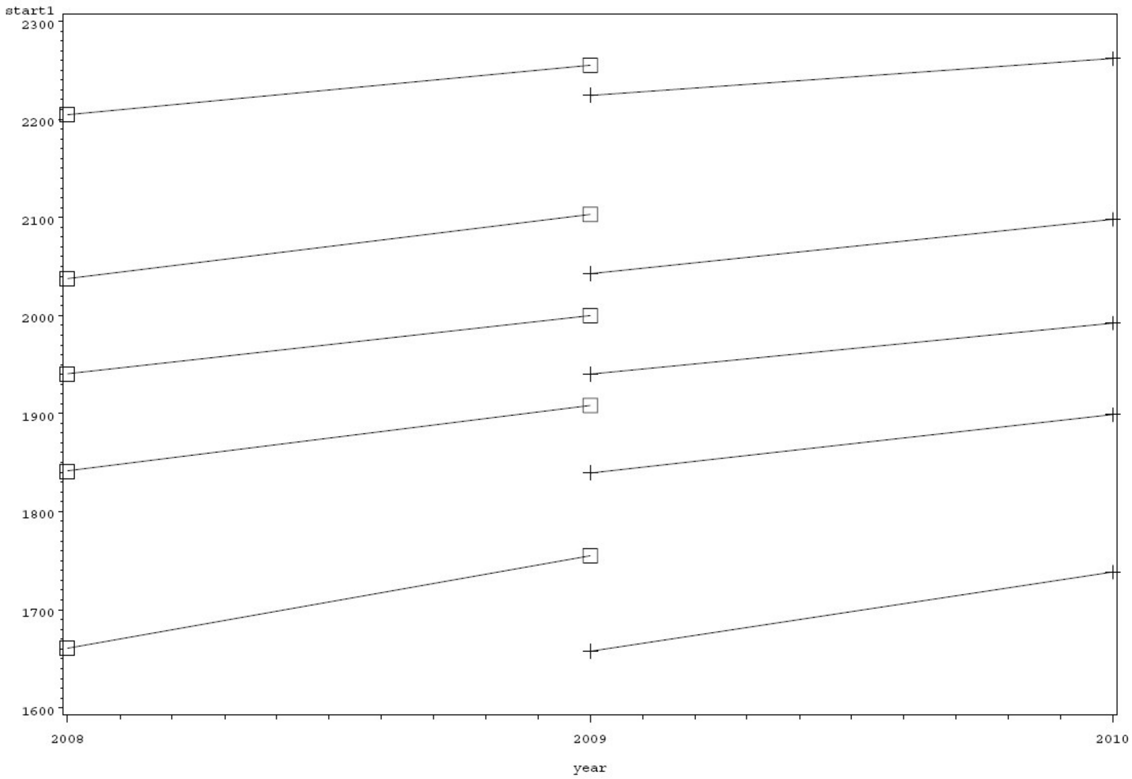
Grade 7 to 8 Transitions for Math 2004 to 2010

Black lines (with squares) are adjacent-year quintile cohorts



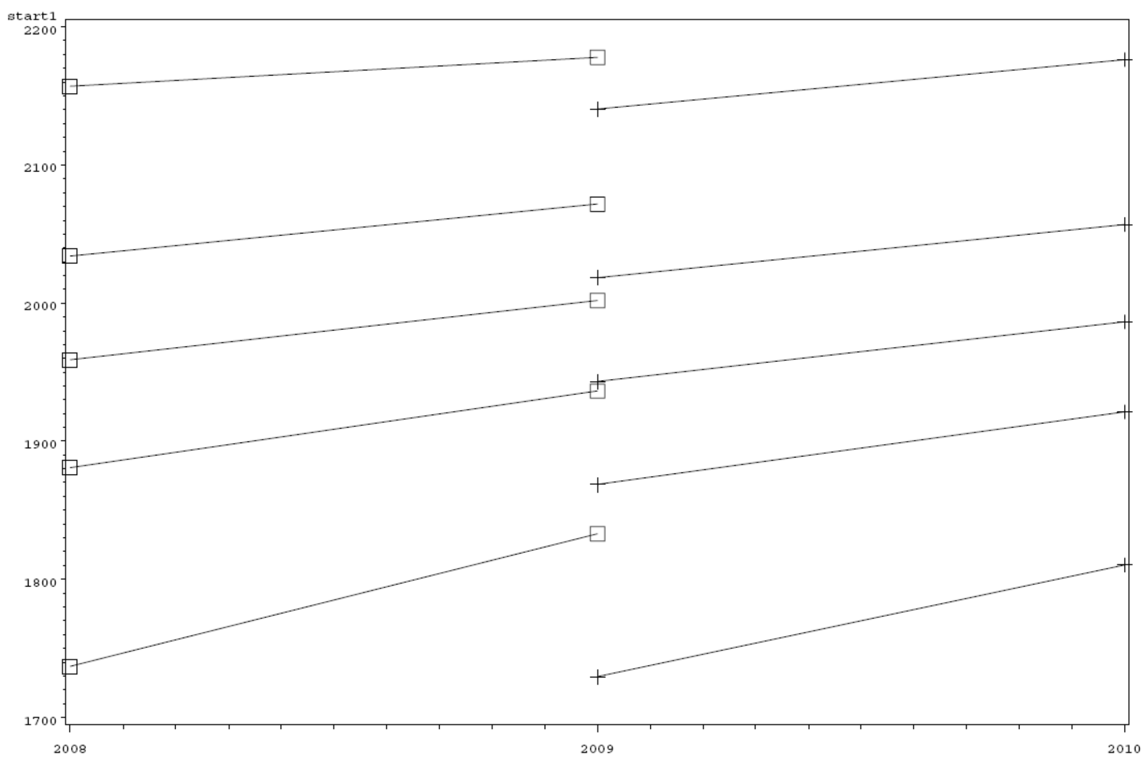
Grade 8 to 9 Transitions for Reading 2004 to 2010

Black lines (with squares) are adjacent-year quintile cohorts



Grade 8 to 9 Transitions for Math 2004 to 2010

Black lines (with squares) are adjacent-year quintile cohorts

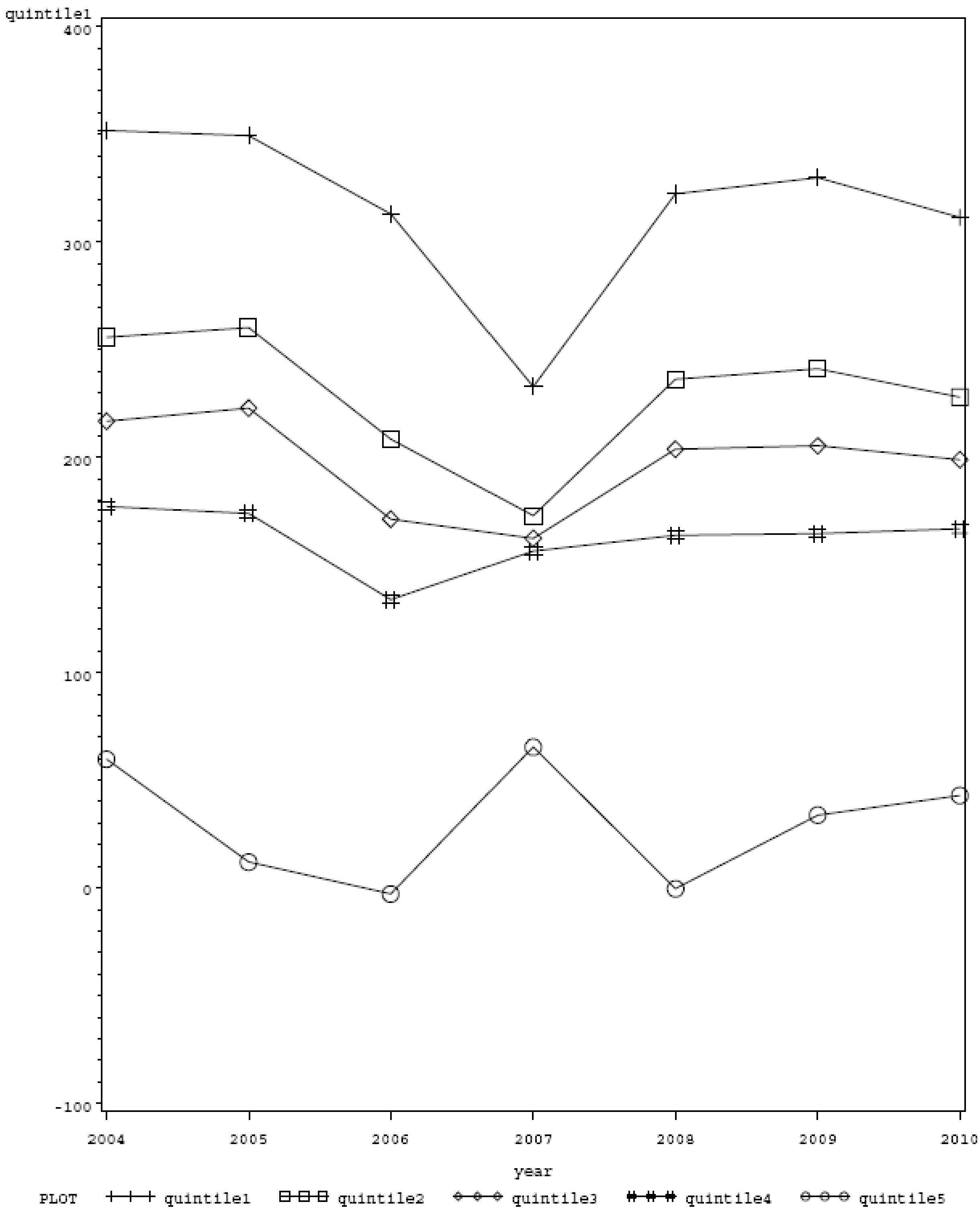


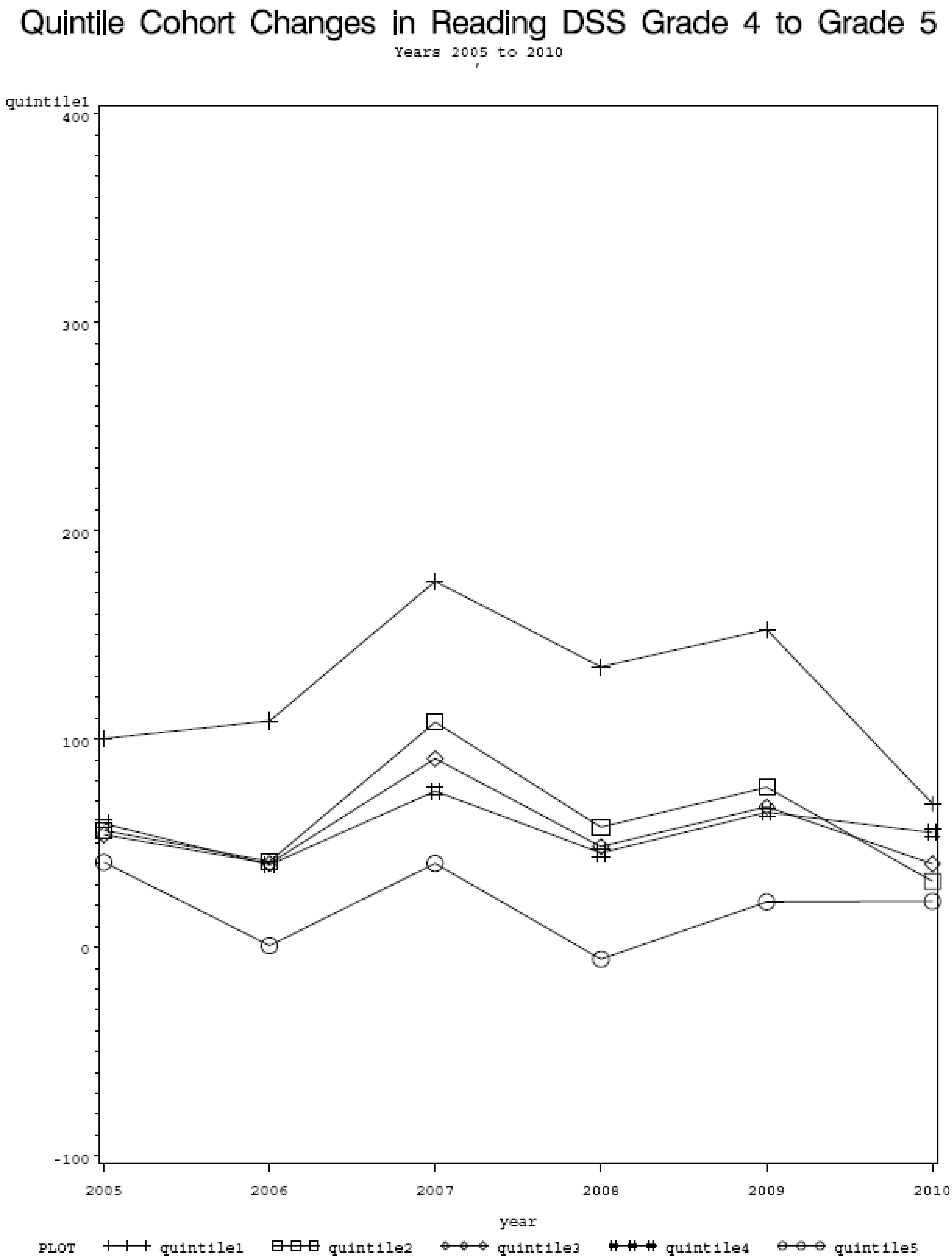
Appendix C

Score Gain Histories 2003 to 2010

Quintile Cohort Changes in Reading DSS Grade 3 to Grade 4

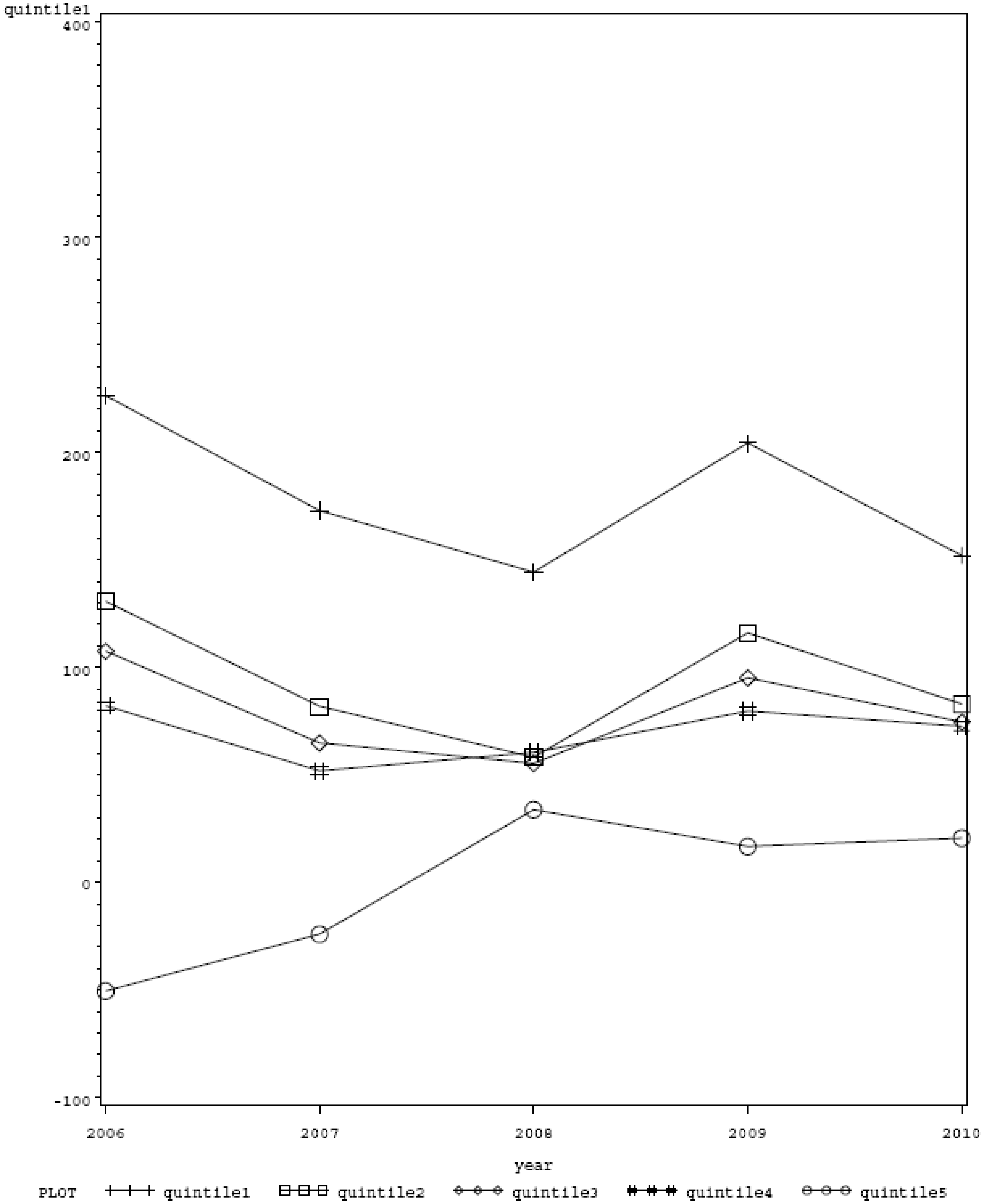
Years 2004 to 2010




Quintile Cohort Changes in Reading DSS Grade 4 to Grade 5
Years 2005 to 2010
quintile1
400
300
200
100
0
-100
2005
2006
2007
2008
2009
2010
year
PLOT
quintile1
quintile2
quintile3
quintile4
quintile5

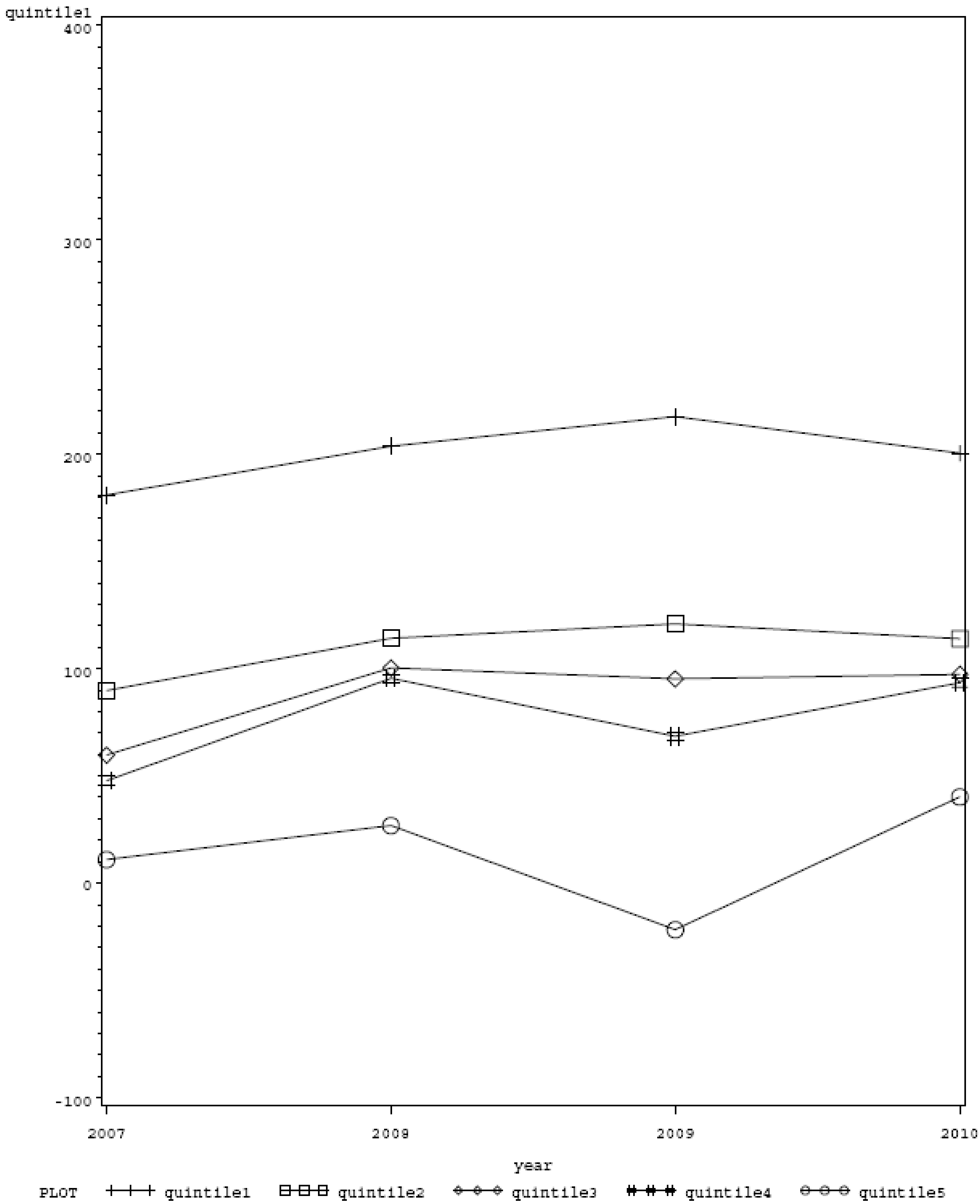
Quintile Cohort Changes in Reading DSS Grade 5 to Grade 6

Years 2006 to 2010



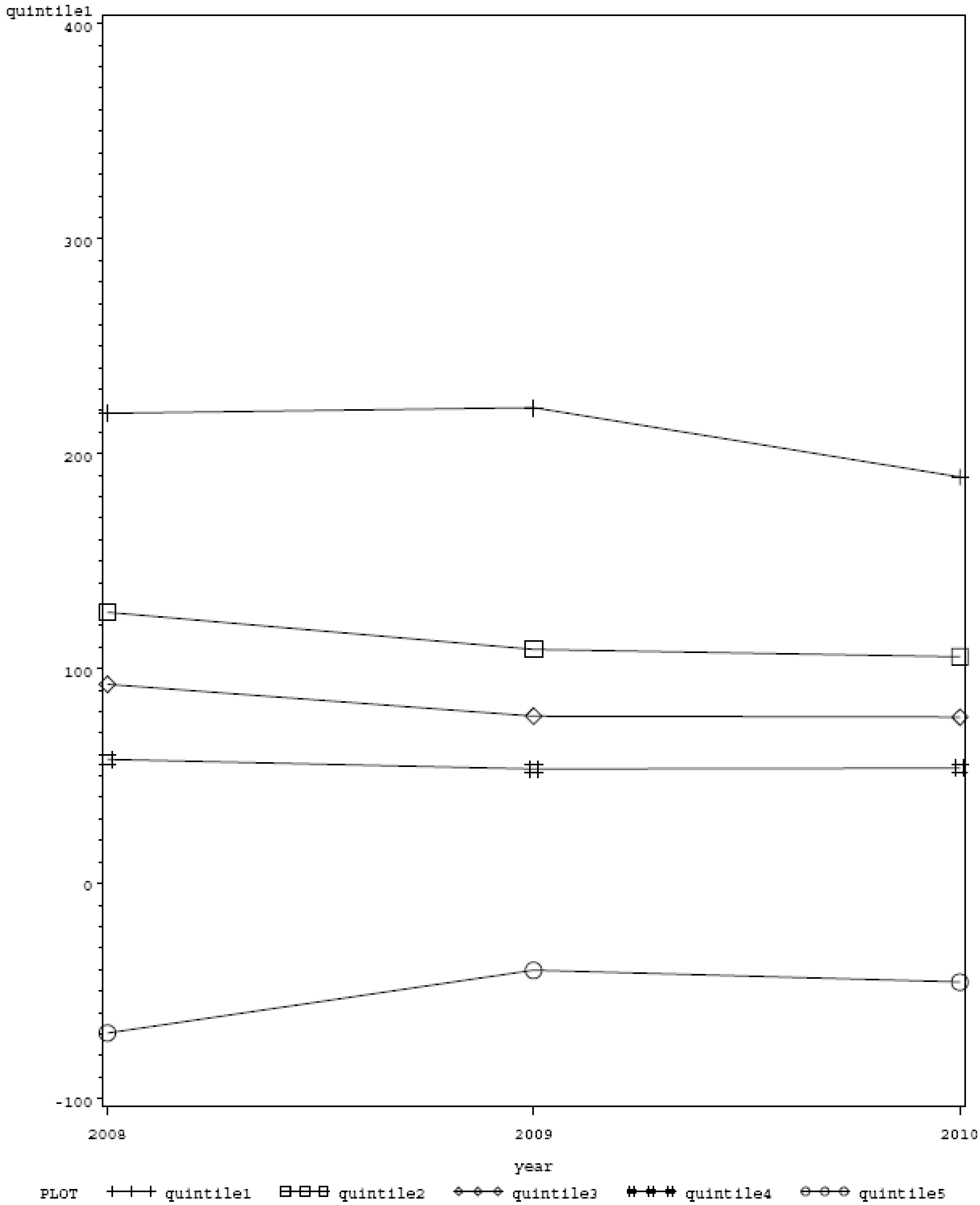
Quintile Cohort Changes in Reading DSS Grade 6 to Grade 7

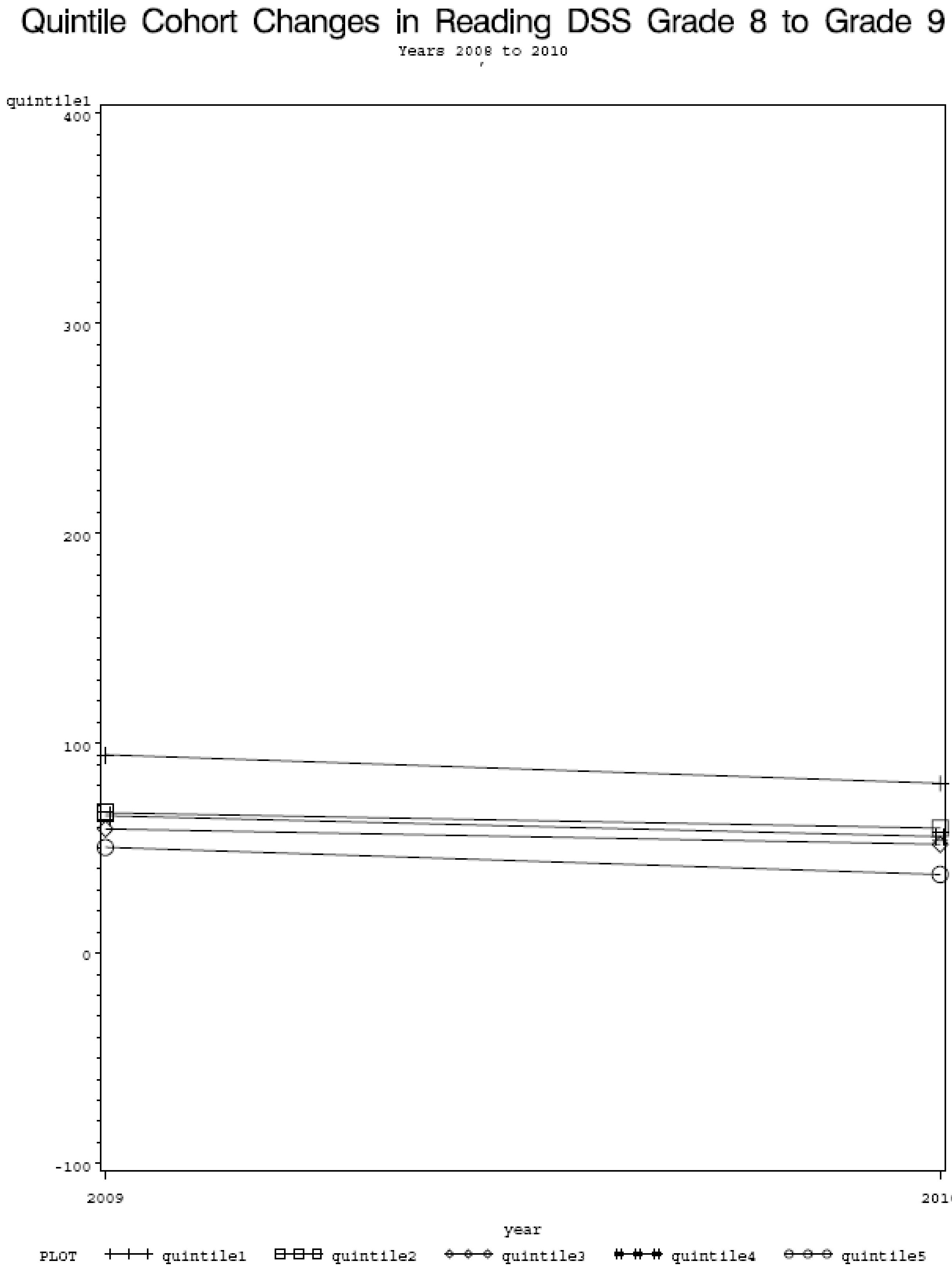
Years 2007 to 2010



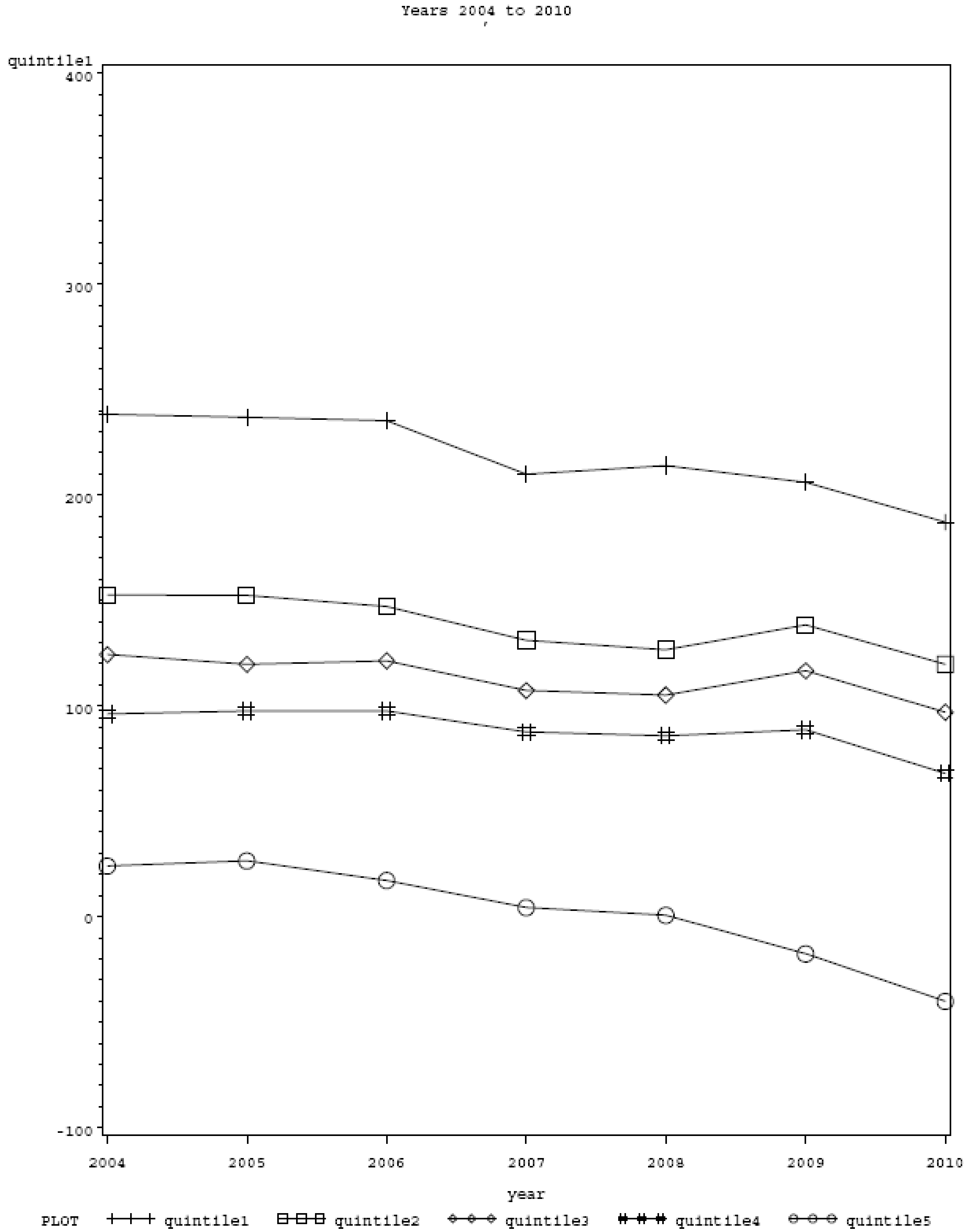
Quintile Cohort Changes in Reading DSS Grade 7 to Grade 8

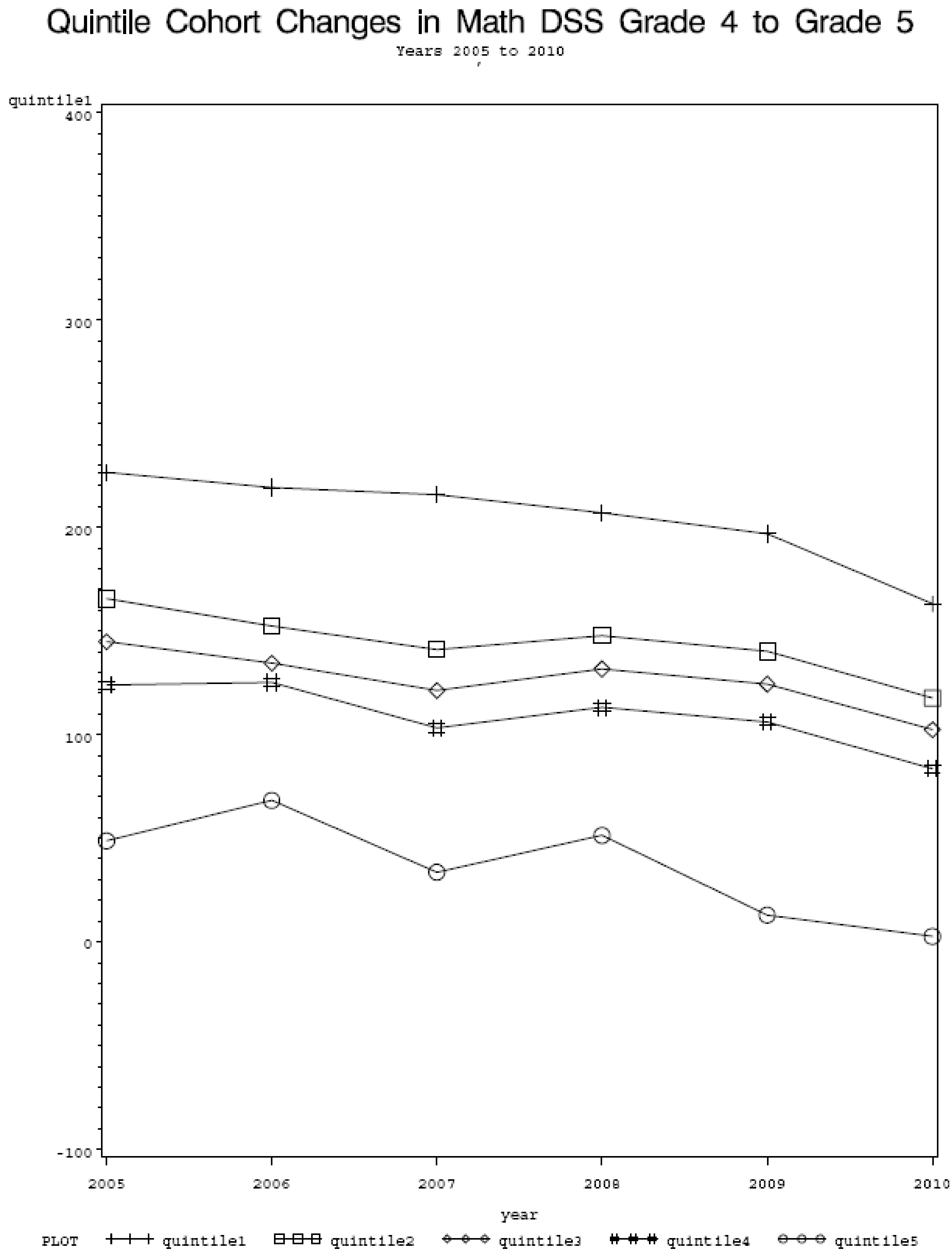
Years 2008 to 2010




Quintile Cohort Changes in Reading DSS Grade 8 to Grade 9
Years 2008 to 2010
quintile1
400
300
200
100
0
-100
2009
2010
year
PLOT
quintile1
quintile2
quintile3
quintile4
quintile5

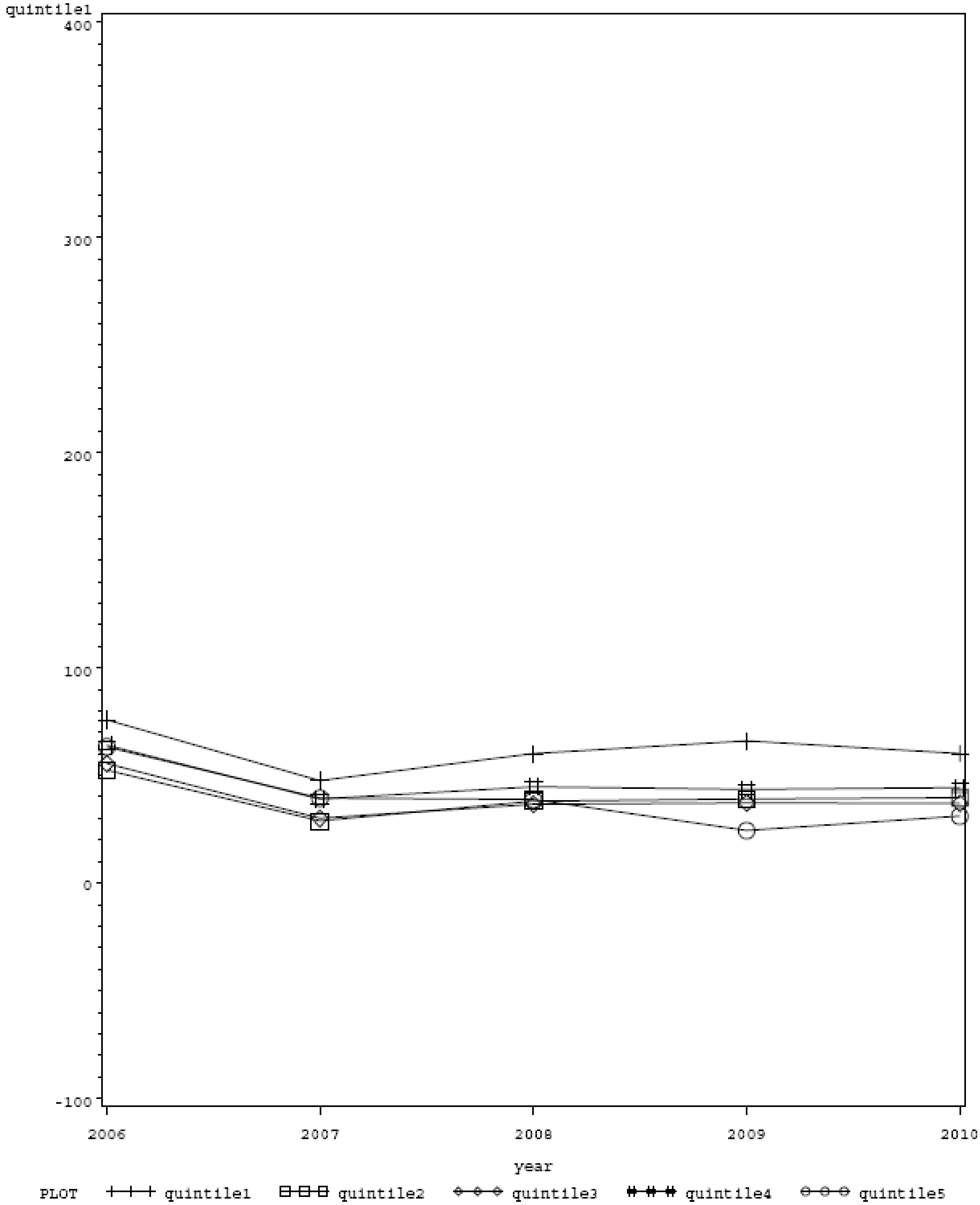
Quintile Cohort Changes in Math DSS Grade 3 to Grade 4




Quintile Cohort Changes in Math DSS Grade 4 to Grade 5
Years 2005 to 2010
quintile1
400
300
200
100
0
-100
2005
2006
2007
2008
2009
2010
year
PLOT
quintile1
quintile2
quintile3
quintile4
quintile5

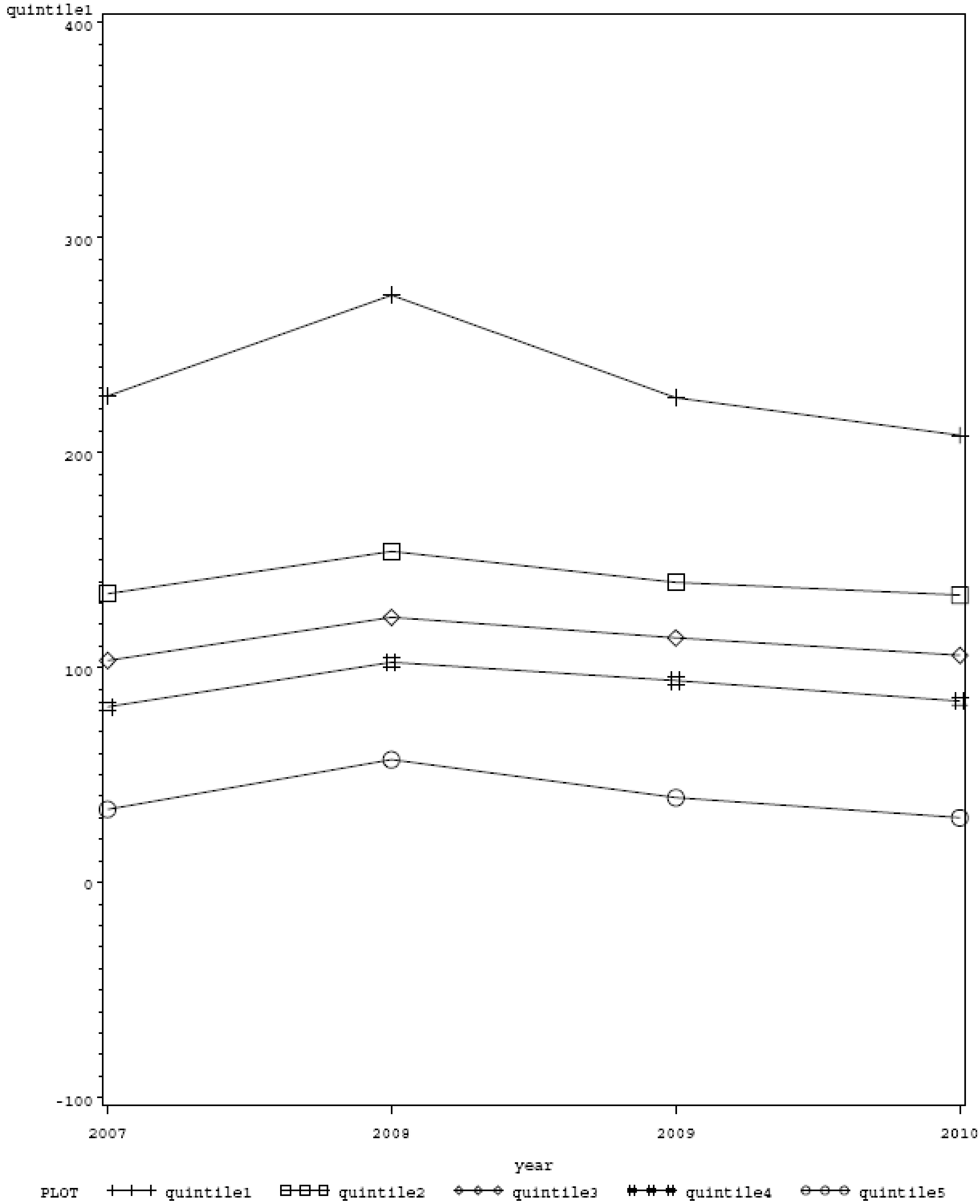
Quintile Cohort Changes in Math DSS Grade 5 to Grade 6

Years 2006 to 2010

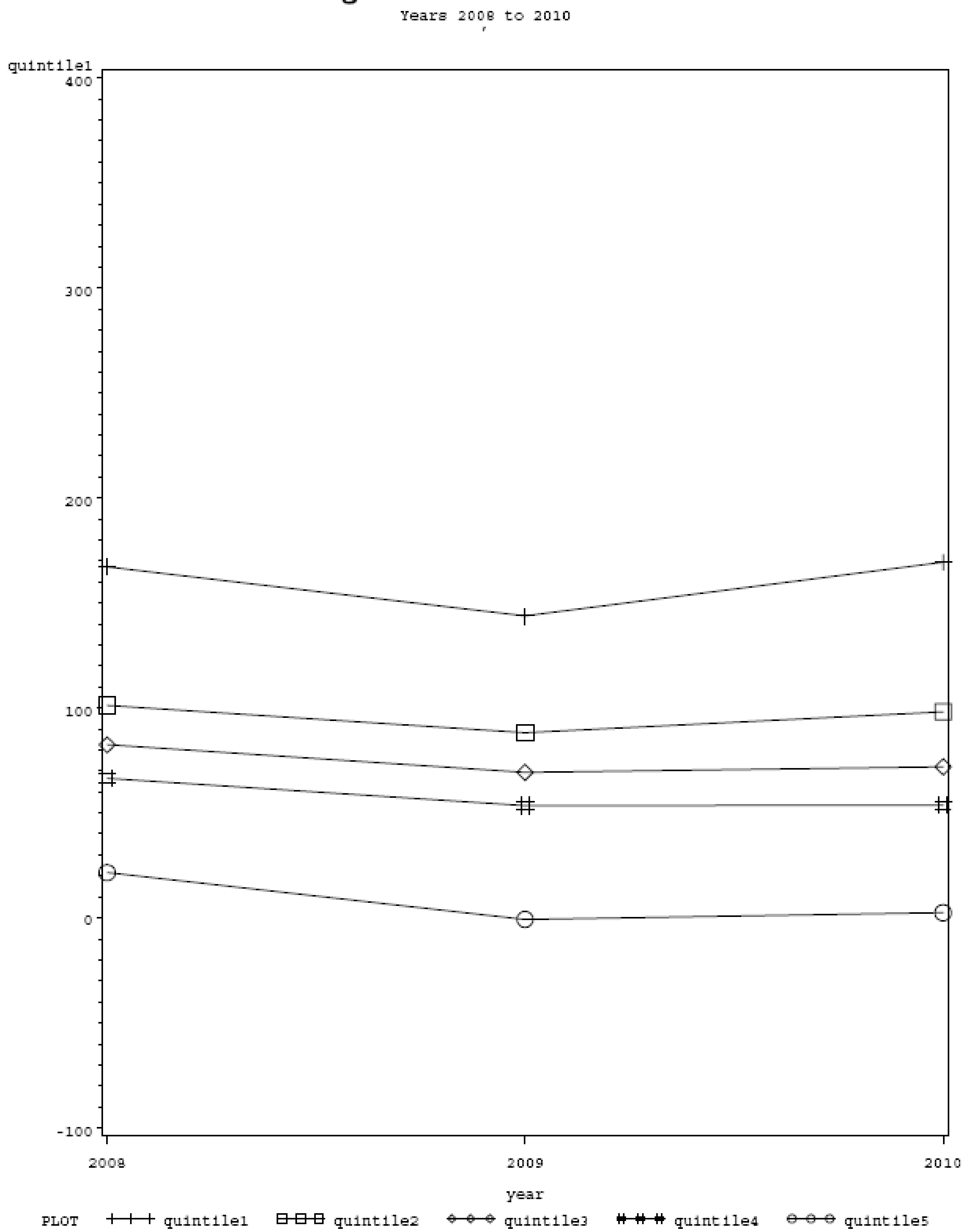


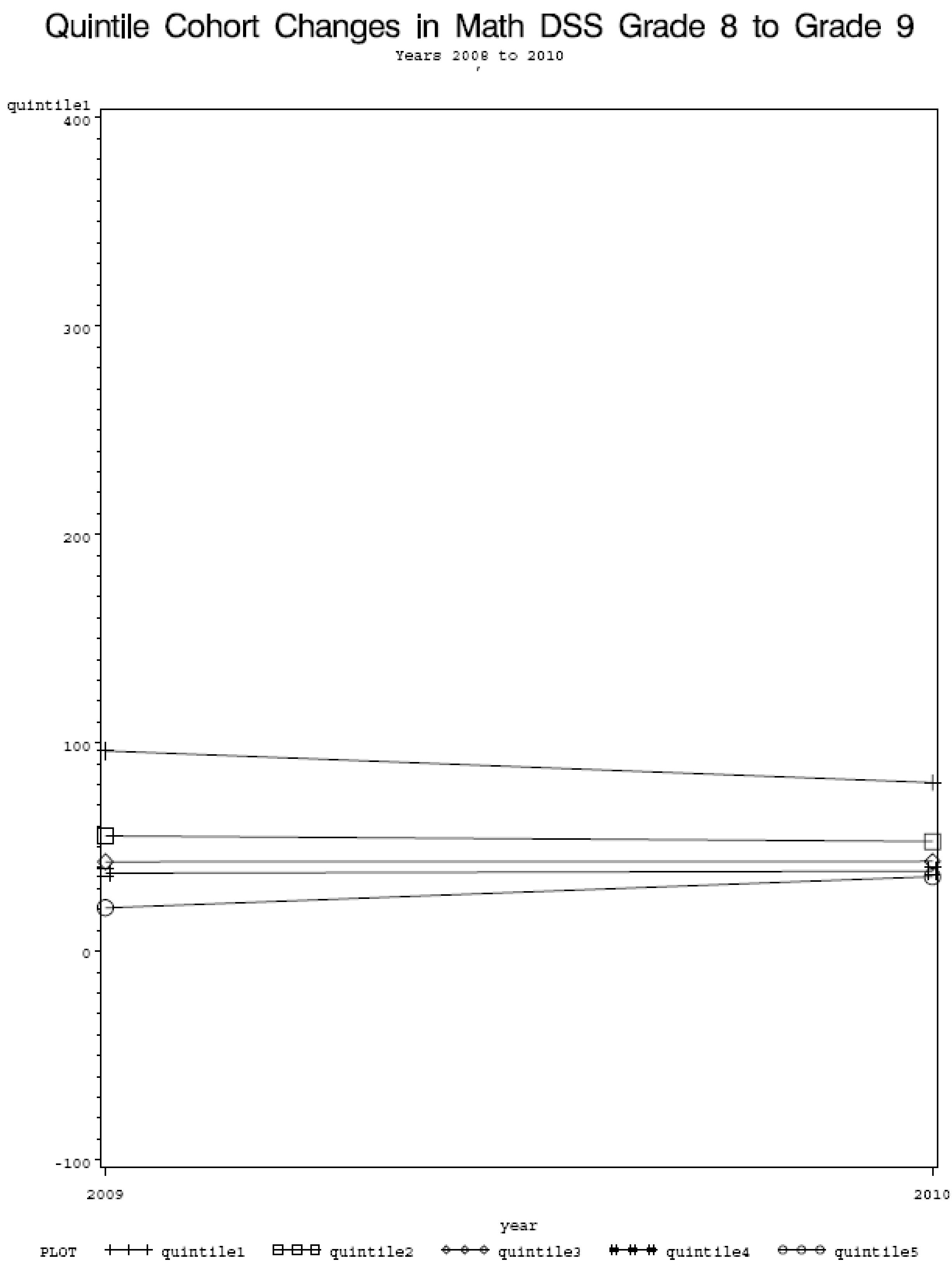
Quintile Cohort Changes in Math DSS Grade 6 to Grade 7

Years 2007 to 2010



Quintile Cohort Changes in mMath DSS Grade 7 to Grade 8




Quintile Cohort Changes in Math DSS Grade 8 to Grade 9
Years 2008 to 2010
quintile1
400
300
200
100
0
-100
2009
2010
year
PLOT
quintile1
quintile2
quintile3
quintile4
quintile5

Appendix D

Relationships between Scores across Grades for Students

Reading

Predictor	Grade 4 DSS R^2	Change in R^2	Total Change in R^2	Points
Grade 3 DSS	0.540			
and Year	0.544	0.003		
and Interaction	0.546	0.003	0.006	20
	Grade 3 to Grade 4 DSS Change			
	R^2	Change in R^2	Total Change in R^2	
Grade 3 DSS	0.288			
and Year	0.293	0.005		
and Interaction	0.297	0.004	0.009	

Predictor	Grade 5 DSS R^2	Change in R^2	Total Change in R^2	Points
Grade 4 DSS	0.576			
and Year	0.580	0.003		
and Interaction	0.582	0.002	0.006	21
	Grade 4 to Grade 5 DSS Change			
	R^2	Change in R^2	Total Change in R^2	
Grade 4 DSS	0.064			
and Year	0.072	0.008		
and Interaction	0.077	0.005	0.012	

Predictor	Grade 6 DSS R^2	Change in R^2	Total Change in R^2	Points
Grade 5 DSS	0.580			
and Year	0.582	0.002		
and Interaction	0.586	0.004	0.006	21
	Grade 5 to Grade 6 DSS Change			
	R^2	Change in R^2	Total Change in R^2	
Grade 5 DSS	0.142			
and Year	0.145	0.004		
and Interaction	0.153	0.008	0.012	

Reading (con't)

Predictor	Grade 7 DSS R^2	Change in R^2	Total Change in R^2	Points
Grade 6 DSS	.590			
and Year	.592	.002		
and Interaction	.593	.001	.003	
	Grade 6 to Grade 7 DSS Change			
	R^2	Change in R^2	Total Change in R^2	
Grade 6 DSS	.182			
and Year	.185	.003		
and Interaction	.188	.003	.006	

Predictor	Grade 8 DSS R^2	Change in R^2	Total Change in R^2	Points
Grade 7 DSS	0.613			
and Year	0.613	0.000		
and Interaction	0.614	0.001	0.001	7
	Grade 7 to Grade 8 DSS Change			
	R^2	Change in R^2	Total Change in R^2	
Grade 7 DSS	0.404			
and Year	0.404	0.000		
and Interaction	0.406	0.001	0.001	

Predictor	Grade 9 DSS R^2	Change in R^2	Total Change in R^2	Points
Grade 8 DSS	0.606			
and Year	0.606	0.000		
and Interaction	0.606	0.000	0.000	5
	Grade 8 to Grade 9 DSS Change			
	R^2	Change in R^2	Total Change in R^2	
Grade 8 DSS	0.021			
and Year	0.022	0.001		
and Interaction	0.022	0.000	0.001	

Mathematics

	Grade 4 DSS			
Predictor	R^2	Change in R^2	Total Change in R^2	Points
Grade 3 DSS	0.597			
and Year	0.598	0.001		
and Interaction	0.598	0.000	0.001	6
	Grade 3 to Grade 4 DSS Change			
	R^2	Change in R^2	Total Change in R^2	
Grade 3 DSS	0.255			
and Year	0.256	0.001		
and Interaction	0.256	0.000	0.001	

	Grade 5 DSS			
Predictor	R^2	Change in R^2	Total Change in R^2	Points
Grade 4 DSS	0.648			
and Year	0.651	0.003		
and Interaction	0.652	0.000	0.003	11
	Grade 4 to Grade 5 DSS Change			
	R^2	Change in R^2	Total Change in R^2	
Grade 5 DSS	0.257			
and Year	0.263	0.006		
and Interaction	0.264	0.001	0.007	

	Grade 6 DSS			
Predictor	R^2	Change in R^2	Total Change in R^2	Points
Grade 5 DSS	0.679			
and Year	0.680	0.001		
and Interaction	0.681	0.000	0.002	9
	Grade 5 to Grade 6 DSS Change			
	R^2	Change in R^2	Total Change in R^2	
Grade 5 DSS	0.015			
and Year	0.019	0.004		
and Interaction	0.020	0.001	0.005	

Mathematics (con't)

Predictor	Grade 7 DSS R^2	Change in R^2	Total Change in R^2	Points
Grade 6 DSS	0.658			
and Year	0.660	0.002		
and Interaction	0.661	0.001	0.003	
	Grade 6 to Grade 7 DSS Change			
	R^2	Change in R^2	Total Change in R^2	
Grade 6 DSS	0.340			
and Year	0.345	0.005		
and Interaction	0.347	0.002	0.008	

Predictor	Grade 8 DSS R^2	Change in R^2	Total Change in R^2	Points
Grade 7 DSS	0.709			
and Year	0.711	0.002		
and Interaction	0.712	0.000	0.002	7
	Grade 7 to Grade 8 DSS Change			
	R^2	Change in R^2	Total Change in R^2	
Grade 7 DSS	0.384			
and Year	0.388	0.004		
and Interaction	0.389	0.001	0.001	

Predictor	Grade 9 DSS R^2	Change in R^2	Total Change in R^2	Points
Grade 8 DSS	0.708			
and Year	0.708	0.000		
and Interaction	0.709	0.001	0.001	5
	Grade 8 to Grade 9 DSS Change			
	R^2	Change in R^2	Total Change in R^2	
Grade 8 DSS	0.099			
and Year	0.100	0.000		
and Interaction	0.102	0.003	0.003	

Appendix E

Relationships between Scores across Grades for Schools

Reading

Predictor	Grade 4 DSS R2	Change in R2	Total Change in R2	Points
Grade 3 DSS	0.839			
and Year	0.859	0.020		
and Interaction	0.863	0.004	0.024	16
	Grade 3 to Grade 4 DSS Change			
	R2	Change in R2	Total Change in R2	
Grade 3 DSS	0.173			
and Year	0.276	0.103		
and Interaction	0.295	0.019	0.121	

Predictor	Grade 5 DSS R2	Change in R2	Total Change in R2	Points
Grade 4 DSS	0.870			
and Year	0.901	0.031		
and Interaction	0.904	0.003	0.034	21
	Grade 4 to Grade 5 DSS Change			
	R2	Change in R2	Total Change in R2	
Grade 4 DSS	0.000			
and Year	0.238	0.238		
and Interaction	0.261	0.023	0.261	

Predictor	Grade 6 DSS R2	Change in R2	Total Change in R2	Points
Grade 5 DSS	0.884			
and Year	0.899	0.015		
and Interaction	0.904	0.005	0.020	16
	Grade 5 to Grade 6 DSS Change			
	R2	Change in R2	Total Change in R2	
Grade 5 DSS	0.007			
and Year	0.138	0.130		
and Interaction	0.177	0.039	0.170	

Reading (con't)

	Grade 7 DSS			
Predictor	R2	Change in R2	Total Change in R2	Points
Grade 6 DSS	0.908			
and Year	0.918	0.029		
and Interaction	0.921	0.003	0.032	
	Grade 6 to Grade 7 DSS Change			
	R2	Change in R2	Total Change in R2	
Grade 6 DSS	0.135			
and Year	0.232	0.097		
and Interaction	0.261	0.029	0.068	

	Grade 8 DSS			
Predictor	R2	Change in R2	Total Change in R2	Points
Grade 7 DSS	0.929			
and Year	0.932	0.003		
and Interaction	0.934	0.002	0.005	6
	Grade 7 to Grade 8 DSS Change			
	R2	Change in R2	Total Change in R2	
Grade 7 DSS	0.423			
and Year	0.448	0.025		
and Interaction	0.461	0.013	0.038	

	Grade 9 DSS			
Predictor	R2	Change in R2	Total Change in R2	Points
Grade 8 DSS	0.940			
and Year	0.943	0.003		
and Interaction	0.944	0.000	0.004	6
	Grade 8 to Grade 9 DSS Change			
	R2	Change in R2	Total Change in R2	
Grade 8 DSS	0.207			
and Year	0.251	0.044		
and Interaction	0.256	0.005	0.049	

Mathematics

Predictor	R^2	Change in R^2	Total Change in R^2	Points
	Grade 4 DSS			
Grade 3 DSS	0.804			
and Year	0.811	0.007		
and Interaction	0.813	0.002	0.009	8
	Grade 3 to Grade 4 DSS Change			
	R^2	Change in R^2	Total Change in R^2	
Grade 3 DSS	0.266			
and Year	0.292	0.026		
and Interaction	0.298	0.006	0.032	

Predictor	R^2	Change in R^2	Total Change in R^2	Points
	Grade 5 DSS			
Grade 4 DSS	0.820			
and Year	0.848	0.028		
and Interaction	0.848	0.000	0.028	14
	Grade 4 to Grade 5 DSS Change			
	R^2	Change in R^2	Total Change in R^2	
Grade 4DSS	0.194			
and Year	0.318	0.124		
and Interaction	0.320	0.002	0.126	

Predictor	R^2	Change in R^2	Total Change in R^2	Points
	Grade 6 DSS			
Grade 5 DSS	0.836			
and Year	0.842	0.006		
and Interaction	0.844	0.001	0.007	8
	Grade 5 to Grade 6 DSS Change			
	R^2	Change in R^2	Total Change in R^2	
Grade 5 DSS	0.046			
and Year	0.081	0.035		
and Interaction	0.089	0.008	0.043	

Mathematics (con't)

Predictor	R^2	Change in R^2	Total Change in R^2	Points
		Grade 7 DSS		
Grade 6 DSS	0.875			
and Year	0.889	0.014		
and Interaction	0.891	0.002	0.016	
		Grade 6 to Grade 7 DSS Change		
	R^2	Change in R^2	Total Change in R^2	
Grade 6 DSS	0.331			
and Year	0.404	0.013		
and Interaction	0.415	0.011	0.0.24	

Predictor	R^2	Change in R^2	Total Change in R^2	Points
		Grade 8 DSS		
Grade 7 DSS	0.929			
and Year	0.937	0.009		
and Interaction	0.938	0.000	0.009	6
		Grade 7 to Grade 8 DSS Change		
	R^2	Change in R^2	Total Change in R^2	
Grade 7 DSS	0.456			
and Year	0.523	0.067		
and Interaction	0.526	0.003	0.070	

Predictor	R^2	Change in R^2	Total Change in R^2	Points
		Grade 9 DSS		
Grade 8 DSS	0.933			
and Year	0.934	0.001		
and Interaction	0.938	0.005	0.005	5
		Grade 8 to Grade 9 DSS Change		
	R^2	Change in R^2	Total Change in R^2	
Grade 8 DSS	0.043			
and Year	0.053	0.011		
and Interaction	0.120	0.066	0.077	

Appendix F

Relationships between Scores across Grades for Districts

Reading

Predictor	R^2	Change in R^2	Total Change in R^2	Points
	Grade4 DSS			
Grade 3 DSS	0.730			
and Year	0.821	0.092		
and Interaction	0.830	0.009	0.100	20
	Grade 3 to Grade 9 DSS Change			
	R^2	Change in R^2	Total Change in R^2	
Grade 3 DSS	0.142			
and Year	0.432	0.291		
and Interaction	0.461	0.028	0.319	

Predictor	R^2	Change in R^2	Total Change in R^2	Points
	Grade 5 DSS			
Grade 4 DSS	0.780			
and Year	0.886	0.106		
and Interaction	0.890	0.004	0.110	23
	Grade 4 to Grade 5 DSS Change			
	R^2	Change in R^2	Total Change in R^2	
Grade 4 DSS	0.002			
and Year	0.481	0.480		
and Interaction	0.499	0.018	0.498	

Predictor	R^2	Change in R^2	Total Change in R^2	Points
	Grade 6 DSS			
Grade 5 DSS	0.805			
and Year	0.844	0.039		
and Interaction	0.850	0.006	0.045	15
	Grade 5 to Grade 6 DSS Change			
	R^2	Change in R^2	Total Change in R^2	
Grade 5 DSS	0.038			
and Year	0.228	0.191		
and Interaction	0.259	0.031	0.222	

Reading (con't)

Grade 7 DSS				
Predictor	R^2	Change in R^2	Total Change in R^2	Points
Grade 6 DSS	0.848			
and Year	0.887	0.039		
and Interaction	0.890	0.003	0.041	
Grade 6 to Grade 7 DSS Change				
	R^2	Change in R^2	Total Change in R^2	
Grade 6 DSS	0.203			
and Year	0.406	0.203		
and Interaction	0.418	0.012	0.215	

Grade 8 DSS				
Predictor	R^2	Change in R^2	Total Change in R^2	Points
Grade 7 DSS	0.916			
and Year	0.931	0.015		
and Interaction	0.931	0.001	0.015	6
Grade 7 to Grade 8 DSS Change				
	R^2	Change in R^2	Total Change in R^2	
Grade 7 DSS	0.437			
and Year	0.535	0.097		
and Interaction	0.540	0.005	0.005	

Grade 9 DSS				
Predictor	R^2	Change in R^2	Total Change in R^2	Points
Grade 8 DSS	0.930			
and Year	0.934	0.004		
and Interaction	0.934	0.000	0.004	4
Grade 8 to Grade 9 DSS Change				
	R^2	Change in R^2	Total Change in R^2	
Grade 8 DSS	0.239			
and Year	0.278	0.039		
and Interaction	0.280	0.002	0.041	

Mathematics

Grade 4 DSS				
Predictor	R^2	Change in R^2	Total Change in R^2	Points
Grade 3 DSS	0.769			
and Year	0.783	0.014		
and Interaction	0.785	0.001	0.016	7
Grade 3 to Grade 4 DSS Change				
	R^2	Change in R^2	Total Change in R^2	
Grade 3 DSS	0.224			
and Year	0.271	0.048		
and Interaction	0.276	0.005	0.052	

Grade 5 DSS				
Predictor	R^2	Change in R^2	Total Change in R^2	Points
Grade 4 DSS	0.757			
and Year	0.810	0.053		
and Interaction	0.811	0.001	0.054	12
Grade 4 to Grade 5 DSS Change				
	R^2	Change in R^2	Total Change in R^2	
Grade 4 DSS	0.190			
and Year	0.365	0.176		
and Interaction	0.369	0.004	0.180	

Grade 6 DSS				
Predictor	R^2	Change in R^2	Total Change in R^2	Points
Grade 5 DSS	0.770			
and Year	0.783	0.013		
and Interaction	0.786	0.003	0.016	8
Grade 5 to Grade 6 DSS Change				
	R^2	Change in R^2	Total Change in R^2	
Grade 5 DSS	0.000			
and Year	0.056	0.056		
and Interaction	0.069	0.013	0.069	

Mathematics (con't)

Predictor	R^2	Change in R^2	Total Change in R^2	Points
		Grade 7 DSS		
Grade 6 DSS	0.828			
and Year	0.872	0.042		
and Interaction	0.874	0.002	0.044	
		Grade 6 to Grade 7 DSS Change		
	R^2	Change in R^2	Total Change in R^2	
Grade 6 DSS	0.319			
and Year	0.493	0.174		
and Interaction	0.501	0.008	0.437	

Predictor	R^2	Change in R^2	Total Change in R^2	Points
		Grade 8 DSS		
Grade 7 DSS	0.870			
and Year	0.911	0.041		
and Interaction	0.911	0.000	0.041	8
		Grade 7 to Grade 8 DSS Change		
	R^2	Change in R^2	Total Change in R^2	
Grade 7 DSS	0.441			
and Year	0.618	0.177		
and Interaction	0.619	0.001	0.001	

Predictor	R^2	Change in R^2	Total Change in R^2	Points
		Grade 9 DSS		
Grade 8 DSS	0.929			
and Year	0.929	0.000		
and Interaction	0.934	0.005	0.005	3
		Grade 8 to Grade 9 DSS Change		
	R^2	Change in R^2	Total Change in R^2	
Grade 8 DSS	0.033			
and Year	0.038	0.005		
and Interaction	0.103	0.065	0.070	