9-12-2006

Exploring the Efficacy of Pre-Equating a Large Scale Criterion-Referenced Assessment with Respect to Measurement Equivalence

Christopher Stephen Domaleski
domaleski@usa.net

ACCEPTANCE

This dissertation, EXPLORING THE EFFICACY OF PRE-EQUATING A LARGE SCALE CRITERION-REFERENCED ASSESSMENT WITH RESPECT TO MEASUREMENT EQUIVALENCE, by CHRISTOPHER S. DOMALESKI, was prepared under the direction of the candidate's Dissertation Advisory Committee. It is accepted by the committee members in partial fulfillment of the requirements for the degree Doctor of Philosophy in the College of Education, Georgia State University.

The Dissertation Advisory Committee and the student's Department Chair, as representatives of the faculty, certify that this dissertation has met all standards of excellence and scholarship as determined by the faculty. The Dean of the College of Education concurs.

T. Chris Oshima, Ph.D.
Committee Chair

William L. Curlette, Ph.D.
Committee Member

Toshi Kii, Ph.D.
Committee Member

Malina K. Monaco, Ph.D.
Committee Member

John H. Neel, Ph.D.
Committee Member

Carolyn Furlow, Ph.D.
Committee Member

Date

Sheryl Gowen, Ph.D.
Chair, Department of Educational Policy Studies

Ronald P. Colarusso, Ed.D.
Dean

AUTHOR'S STATEMENT

By presenting this dissertation as a partial fulfillment of the requirements for the advanced degree from Georgia State University, I agree that the library of Georgia State University shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to quote, to copy from, or to publish this dissertation may be granted by the professor under whose direction it was written, by the College of Education's director of graduate studies and research, or by me. Such quoting, copying, or publishing must be solely for scholarly purposes and will not involve potential financial gain. It is understood that any copying from or publication of this dissertation which involves potential financial gain will not be allowed without my written permission.

_____

Christopher S. Domaleski

NOTICE TO BORROWERS

All dissertations deposited in the Georgia State University library must be used in accordance with the stipulations prescribed by the author in the preceding statement. The author of this dissertation is:

Christopher S. Domaleski
3660 Sugar Creek Lane
Conyers, Georgia, 30094

The director of this dissertation is:

Dr. T. Chris Oshima
Department of Educational Policy Studies
College of Education
Georgia State University
Atlanta, GA 30303-3083

VITA

Christopher Stephen Domaleski

ADDRESS: 3660 Sugar Creek Lane
Conyers, Georgia 30094

EDUCATION:

Ph.D. 2006	Georgia State University Educational Policy Studies
M.Ed. 1994	University of Georgia Student Personnel in Higher Education
B.S. 1991	North Georgia College Political Science

PROFESSIONAL EXPERIENCE:

2003-Present	Program Manager for Research and Development Georgia Department of Education, Atlanta, GA
2002-2003	Assessment Specialist Georgia Department of Education, Atlanta, GA
2000-2001	Director of Student Life Georgia Perimeter College, Dunwoody, GA
1998-2000	Assistant Director of Student Life Centre College, Danville, KY
1994-1998	Area Coordinator Emory and Henry College, Emory, VA

SELECTED PRESENTATIONS AND PUBLICATIONS:

Bishop, N. S., Sharairi, S., Swift, D., Lei, P., & Domaleski, C. S. (2006, April). *Comparing Growth and AYP Results Over Multiple Years Using Different IRT Scale Transformation Procedures.* Paper presented at the annual meeting of the National Council on Measurement and Education, San Francisco, CA.

Bishop, N. S., Sharairi, S., Liu, Y., Dabney, M., Komatsu, L., & Domaleski, C. S. (2004, April). *Comparing equating results over multiple years from dichotomous and polytomous IRT models for a test composed of testlets.* Paper presented at the annual meeting of the National Council on Measurement and Education, San Diego, CA.

Bishop, N. S., Swift, D., Sharairi, S., Lei, P., & Domaleski, C. S. (2006, April). *Alternative Procedures for Identifying Non-Reading Test Items with High Reading Loads.* Paper presented at the annual meeting of the National Council on Measurement and Education, San Francisco, CA.

Domaleski, C. S. (2004, September). *Measuring Success in Third Grade Reading*. Presentation at the annual Georgia Curriculum Director's Conference, Athens, GA.

Domaleski, C. S. (2005, October). *Using CRCT Data to Understand Social Studies Achievement and Inform Instruction.* Presentation at the annual Georgia Council for the Social Studies Conference, Athens, GA.

Neel, J. H., Monaco, M. K., Domaleski, C. S., & Stephens-Bonty, T. (2001, November). *Levene's Test Revisited.* Paper presented at the annual meeting of the Georgia Educational Research Association, Atlanta, GA.

Oshima, T. C., Raju, N. S., & Domaleski, C. S. (2006, April). *Conditional DIF and DTF*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Oshima, T. C. & Domaleski, C. S. (in press). Academic Performance Gap Due to Birthdate in Kindergarten Through Eighth Grade. *Journal of Educational Research*.

ABSTRACT

EXPLORING THE EFFICACY OF PRE-EQUATING A LARGE SCALE CRITERION-REFERENCED ASSESSMENT WITH RESPECT TO MEASUREMENT EQUIVALENCE

by

Christopher S. Domaleski

This investigation examined the practice of relying on field test item calibrations in advance of the operational administration of a large scale assessment for purposes of equating and scaling. Often termed "pre-equating," the effectiveness of this method is explored for a statewide, high-stakes assessment in grades three, five, and seven for the content areas of language arts, mathematics, and social studies.

Pre-equated scaling was based on item calibrations using the Rasch model from an off-grade field test event in which students tested were one grade higher than the target population. These calibrations were compared to those obtained from post-equating, which used the full statewide population of examinees. Item difficulty estimates and Test Characteristic Curves (TCC) were compared for each approach and found to be similar. The Root Mean Square Error (RMSE) of the theta estimates for each approach ranged from .02 to .12. Moreover, classification accuracy for the pre-equated approach was generally high compared to results from post-equating. Only 3 of the 9 tests examined showed differences in the percent of students classified as passing; errors ranged from 1.7 percent to 3 percent.

Measurement equivalence between the field test and operational assessment was also explored using the Differential Functioning of Items and Tests (DFIT)

framework. Overall, about 20 to 40 percent of the items on each assessment exhibited statistically significant Differential Item Functioning (DIF). Differential Test Functioning (DTF) was significant for fully 7 tests. There was a positive relationship between the magnitude of DTF and degree of incongruence between pre-equating and post-equating.

Item calibrations, score consistency, and measurement equivalence were also explored for a test calibrated with the one, two, and three parameter logistic model, using the TCC equating method. Measurement equivalence and score table incongruence was found to be slightly more pronounced with this approach.

It was hypothesized that differences between the field test and operational tests resulted from 1) recency of instruction 2) cognitive growth and 3) motivation factors. Additional research related to these factors is suggested.

EXPLORING THE EFFICACY OF PRE-EQUATING A LARGE SCALE CRITERION-REFERENCED ASSESSMENT WITH RESPECT TO MEASUREMENT EQUIVALENCE

by
Christopher S. Domaleski

A Dissertation

Presented in Partial Fulfillment of Requirements for the
Degree of
Doctor of Philosophy
in
Educational Policy Studies
in
the Department of Educational Policy Studies
in
the College of Education
Georgia State University

Atlanta, Georgia
2006

TABLE OF CONTENTS

Page

Acknowledgements ……..……..iii
List of Tables…….. v
List of Figures……vi
Abbreviations……..vii

Chapter
1 THE PROBLEM…….1
Research Questions…….3
Significance ..……..4

2 REVIEW OF THE LITERATURE……..... 5
Equating……..…… 5
Equating Designs…… 9
Pre-Equating……... 10
Measurement Equivalence……..12

3 METHODS…… 17
Instrumentation…….. 17
Procedures.....…….…..18

4 RESULTS……...……25
Phase I…….……....25
Phase II……41

5 DISCUSSION……45
Measurement Equivalence……..45
Score Table Comparison and Classification Accuracy……...49
Items and Domains…… 53
Phase II Results…….. 54
Conclusions……55

References…….57
Appendix……61

ACKNOWLEDGEMENTS

Although the phrase may be a bit hackneyed, I have certainly "stood on the shoulders of giants" with respect to this dissertation. I have benefited greatly from the instruction, guidance, and encouragement of many people.

First, I wish to express my most sincere appreciation to the chair of my dissertation committee, Dr. Chris Oshima. Dr. Oshima is one of the most rare and wonderful people I have ever encountered. As her student in Item Response Theory classes and later seeking her mentorship throughout the preparation of this dissertation, I have never ceased to be impressed with her brilliant mind, which is matched only by her gracious nature. Throughout this dissertation she was unfailingly patient and kind, while never acquiescing on standards for high quality work. This dissertation is greatly improved by her ideas and edits and I literally could not have written this without her.

I would also like to thank the other members of my dissertation committee. Very early in my graduate studies, Dr. Bill Curlette was kind enough to offer me an experience working with him the Educational Research Bureau where I learned more than I thought possible. His patient mentorship has had a great influence on me. I am also grateful to Dr. Carolyn Furlow for her excellent instruction, support, and assistance. My appreciation is also extended to Dr. Toshi Kii for his kindness and for never hesitating to help me out. Additionally, I owe a great deal of gratitude to Dr. John Neel. He is a first-rate instructor and he has always been there for me as I sought him out frequently for advice and

guidance. He is the exemplar of a scholar and educator and a big reason I persevered as a doctoral student.

Dr. Monaco was my first statistics instructor and I credit (or blame) her for instilling in me the spark of interest to pursue a career in educational research. Later as we worked as colleagues at the Department of Education, her influence helped me not only to be a better scholar and researcher, but to be a better person. She is an extraordinarily bright and caring person and I thank her for her mentorship and friendship.

I wish to thank all of my colleagues at the Georgia Department of Education in the Testing Division for their support and guidance. In particular, Dr. Sharron Hunt, Dr. Jeff Barker, and Melissa Fincher have been very helpful and encouraging.

In my role at the Department I had the pleasure of working with Dr. Nambury Raju. He was a truly extraordinary man – a distinguished scholar and an exceedingly congenial gentleman. This dissertation draws heavily on his prolific contributions to psychometric scholarship, chiefly his groundbreaking work with the DFIT framework. I talked with him often about this investigation and benefited immensely from his brilliant guidance. Sadly, Dr. Raju passed away in the fall of 2005. I am among his many friends and colleagues that miss him greatly. This dissertation is dedicated to his memory.

Finally, to my forbearing wife, Kerra, and to my amazing children Christian, Conner, and Kate, I express my deepest love and appreciation. You have been and will always be my inspiration.

LIST OF TABLES

Table		Page
1	Description of Forms and Item Information for 2003 Field-Test	24
2	Summary of RMSE of Item Parameter Estimates	30
3	Summary of RMSE of Theta Estimates	31
4	Summary of Classification Accuracy	33
5	Correlation of Item Difficulty Estimates from Random Sample to Population Parameters	34
6	Frequency/ Percent of Items Demonstrating DIF at .001	35
7	Summary of DTF	36
8	Number and Percent of ELA DIF Items Grade and Domain	37
9	Number and Percent of Mathematics DIF Items Grade and Domain	38
10	Number and Percent of Social Studies DIF Items Grade and Domain	39
11	Frequency and Percent of DIF Items at .001 for Link Set	43
12	Summary of DTF for Link Items	44
13	Summary of Mean Item P-Values	44
14	Summary of Phase II *b* Parameter RMSD	46
15	Summary of Phase II RMSD of Theta Estimates	48
16	Frequency/ Percent of Items Demonstrating DIF at .001	48
17	Differential Test Functioning for One, Two, and Three Parameter Models	49

LIST OF FIGURES

Figure		Page
1	Overview of Study Design…….	27
2	Relationship between DTF and RMSE of Theta Difference……	37
3	Graph of ELA DIF Items by Domain……	41
4	Graph of Mathematics DIF Items by Domain……...	42
5	Graph of Social Studies DIF Items by Domain…….	43
6	Graph of Mean Item P-Values……...	46
7	Example of a Third Grade Item in the ‘Patterns and Relationships’ Domain……..	59

ABBREVIATIONS

1PL	One Parameter Logistic IRT Model
2PL	Two Parameter Logistic IRT Model
3PL	Three Parameter Logistic IRT Model
CDIF	Compensatory Differential Item Functioning
CFA	Confirmatory Factor Analysis
CINEG	Common Item Non-Equivalent Groups
CRCT	Criterion-Referenced Competency Test
CTT	Classic Test Theory
DFIT	Differential Functioning of Items and Tests
DIF	Differential Item Functioning
DTF	Differential Test Functioning
ELA	English Language Arts
GDOE	Georgia Department of Education
ICC	Item Characteristic Curve
IPR	Item Parameter Replication
IRT	Item Response Theory
NCDIF	Non-Compensatory Differential Item Functioning
PL1	Performance Level One for the CRCT ("Does Not Meet Expectations")
PL2	Performance Level Two for the CRCT ("Meets Expectations")

PL3	Performance Level Three for the CRCT (“Exceeds Expectations”)
PLD	Performance Level Descriptors
QCC	Quality Core Curriculum
TCC	Test Characteristic Curve

CHAPTER 1

THE PROBLEM

States are increasingly reliant on information from large-scale assessments for purposes of student and school accountability. Moreover, stakeholders need assessment information quickly to inform decisions related to the requirements of state and/or federal law (e.g. school choice, promotion and retention etc.). In order to meet aggressive scoring and reporting deadlines, test developers may not be able to wait until operational test data are collected to conduct equating and scaling analyses. Therefore, test items are often calibrated prior to the operational administration. Based on these results, number correct (raw score) to scale score conversion tables can be prepared before the test is administered to facilitate rapid reporting. This process is often called *pre-equating*, although some scholars take issue with this potentially confusing term.

There are a number of pre-equating designs available to developers (Lord, 1980; Kolen & Brennan, 1995; Wolfe, 2000). Perhaps the most common technique is to embed a small number of field test items on an operational test. These field test items are not scored, but they are calibrated with the operational items and placed on the same scale. They are then added to a "bank" of items for use on future operational tests. Since the parameters are assumed to be known, a test comprised of such bank items is said to be pre-equated.

Another methodological approach to field testing potential operational items is to conduct a separate testing event, rather than embedding items in an operational test as noted above. Such field test designs may be necessary, for example, when a new test is being developed or any time there is a need for additional pre-equated items prior to an operational administration. For a criterion-referenced test this method presents a unique challenge.

In K-12 criterion-referenced testing students are typically assessed on what they know and can do relative to the standards specified in the curriculum. In Georgia, the Criterion-Referenced Competency Test (CRCT) is one such assessment that purports to measure student proficiency relative to Georgia's Quality Core Curriculum (QCC). The validity of a measure such as the CRCT is reliant on student exposure to QCC based instruction. For this reason, when potential CRCT items are evaluated, the field test sample must include students who have had the opportunity to receive instruction on the QCC for the content area test(s) of interest.

In the fall of 2003 the Georgia Department of Education (GDOE) planned a field test to develop items for an operational test scheduled for the spring of 2004. The goal for this field test was to calibrate new items on the operational scale such that pre-equated operational forms could be developed. In order to insure that students in the field test sample had exposure to all QCC based instruction related to the CRCT, an *off-grade* field test design was implemented. The off-grade field test design simply means that items were administered to students one grade higher than the target grade of the assessment. For example, items being developed for second graders testing in the spring would be

administered to third grade students testing in the fall. This design also insures that the security of the test is maintained.

However, this design raises the question of whether or not the two populations are equivalent. Lack of equivalence could be related to a number of factors including student motivation and recency of instruction. It is hypothesized that the extent to which the off-grade field test model violates assumptions of measurement equivalence will determine the extent to which the pre-equating results can be "trusted." The goal of the present study, then, will be to investigate the relationship between measurement equivalence and pre-equating.

Research Questions

Specifically, the following research questions will be examined:

1) Are the item parameters estimated from the field test (pre-equating) equivalent to those obtained from the operational test (post-equating)?

2) Are the raw to scale score conversion tables obtained in pre-equating equivalent to those calculated in post equating?

3) Will reliance upon pre-equating tables lead to misclassifications in determination of student performance levels compared to post-equating and, if so, to what degree?

4) Will pre-equating results more closely match post-equating results when measurement equivalence for common items at the test level can be established?

5) For what grades, content areas, and item types will differences between pre-equating and post-equating be most pronounced?

6) Will the relationship between pre-equating and post-equating be the same for the one parameter Rasch model compared to the two and three parameter logistic IRT models?

Significance

This investigation is designed to emphasize practical utility. For this reason the analyses will take advantage of actual assessment information collected from an operational, high-stakes testing program. Pre-equating results will be compared to the results obtained from the state-wide, operational event ($n \approx 120{,}000$). Since this administration involves all the students in the population of interest, the operational results will be regarded as "the truth" against which the field test or pre-equating estimates can be compared.

It is expected that the findings of this study will inform practice related to the use of and reliance upon pre-equating. Furthermore, the results may provide information regarding which grades, content areas, and/or item types can be most reliably estimated in pre-equating, versus those that are more subject to fluctuation. The significance of this investigation is underscored by the high-stakes decisions at the student and school level that are based on the precision of tests utilizing similar pre-equating methods.

CHAPTER 2

REVIEW OF THE LITERATURE

Equating

Equating refers to the process of placing item and/or ability estimates on the same scale. By so doing, item parameters or theta estimates for a focal test (X) can be converted to the metric for a reference test (Y). Such a conversion produces two measures that theoretically can be used interchangeably. This practice is important in achievement testing, as it enables one to associate score differences with student performance, rather than as an artifact of a test's changing characteristics.

Early equating procedures from a Classic Test Theory (CTT) framework typically relied on equipercentile or linear approaches (e.g. Angoff, 1971). Equipercentile methods determine equivalent scores by matching respective percentile ranks for each group. If form X is equated to form Y using such an approach, students with the same relative standing in the test distribution will have the same scores. In fact, the distribution of scores on form X will conform to that of form Y in equipercentile equating.

Another CTT approach is linear equating. This approach determines the relationship between forms X and Y by finding the coefficients a and b such that $Y = aX + b$. Such an approach assumes that the performance distribution for forms X and Y differ only in their mean and standard deviation.

However, each of these CTT approaches is of limited utility for current large scale achievement testing because they do not satisfy conditions of symmetry and invariance (Hambleton et al., 1991). When symmetry is established, it is a matter of indifference which test is the reference test and which is the focal test. Invariance holds that the equating procedure is sample independent. Operating within an IRT framework is said to overcome these problems, because ability estimates are not influenced by the distribution from which the estimate was obtained. That is, sample independence is satisfied. In fact, Hambleton et al. assert, "equating of test scores is obviated in an item response theory framework; what must be ensured, however, is that the item and ability parameter values based on two tests are on a common scale" (p. 125). Scaling, not equating, then is the focus of IRT methods.

In this study the "fixed b" or anchored calibration method will be used where Rasch analysis is employed to place tests on the same scale. This is a widely used technique in Rasch analyses (Wolfe, 2000) and is consistent with the method regularly used for the CRCT and many other large scale assessment programs. This process works by fixing the difficulty estimates for a set of anchor items on the focal test, which also appear on a reference test, during the estimation process. Examinee theta estimates and item difficulty estimates are produced such that they are on the same scale as that of the reference test.

The Rasch calibration program Winsteps 3.49 (Linacre & Wright, 2004) can be used to facilitate this process. Winsteps uses a joint maximum likelihood estimation (JMLE) procedure to obtain item and ability estimates. Hambleton et al. define the general form of the JMLE estimation function for N examinees responding to n items:

$$L(u_1, u_2, ..., u_n \mid \theta, b) = \prod_{i=1}^{N} \prod_{j=1}^{n} P_{ij}^{u_{ij}} Q_{ij}^{1-u_{ij}}$$

(2.1)

Where:

u_i = response pattern of examinee *i* to *n* items.

θ = vector of ability parameters

$P_i = P(U_i=1| \theta)$

$Q = 1- P_i$

In the calibration process Winsteps computes theta values based on the anchored items. Then it uses these values to compute item difficulties for the non-anchored values. Iterations continue to increase estimation precision for thetas and item difficulties until convergence is met. Linacre describes the process:

> From the estimation perspective under JMLE, anchored and unanchored items appear exactly alike. The only difference is that anchored values are not changed at the end of each estimation iteration, but unanchored estimates are. JMLE converges when 'observed raw score = expected raw score based on the estimates'. For anchored values, this convergence criterion is never met, but the fit statistics etc. are computed and reported by Winsteps as though it has been met. Convergence is based on the unanchored estimates (personal communication, March 8, 2004).

When the two and three parameter logistic model is used the test characteristic curve (TCC) method attributed to Stocking and Lord (1983) will be used for equating. This method seeks to minimize the squared differences between two test characteristic curves for each value of θ. Hambleton et al. (1991) describe this process as follows:

Given that the true score τ_{xa} of an examinee with ability θ_a on the k common items (*ci*) in test *x* is:

$$\tau_{xa} = \sum_{i=1}^{k} P(\theta_a, b_{xci}, a_{xci}, c_{xci})$$

(2.2)

And the true score τ_{Ya} of an examinee with the same ability on the common items in test *y* is:

$$\tau_{ya} = \sum_{i=1}^{k} P(\theta_a, b_{yci}, a_{yci}, c_{yci})$$

(2.3)

The following transformations are true for the common items:

$$b_{Yci} = \alpha b_{Xci} + \beta$$

(2.4)

$$a_{Yci} = \frac{\alpha_{Xci}}{\alpha}$$

(2.5)

$$c_{Yci} = c_{Xci}$$

(2.6)

Therefore, the constants α and β are determined that minimize the function F for N examinees where:

$$F = \frac{1}{N}\sum_{a=1}^{N}(\tau_{Xa} - \tau_{Ya})^2$$

(2.7)

The TCC method computes this function for a defined number of points on the ability scale, rather than computing a true score for each examinee. Eleven points were used for all TCC analyses in this investigation. Additional details about this method and the iterative procedure required to determine α and β is described in Stocking and Lord (1983).

Equating Designs

In practice, there are several designs that can be employed to obtain the information necessary for equating. Kolen and Brennan (1995) describe three main approaches: random group designs, single group designs, and common-item nonequivalent groups (CINEG).

Random group designs are appealing for reasons of simplicity and security. Each examinee takes only one form of the test and it is not necessary for items to overlap between or among forms. In practice it is common to 'spiral' test forms within the sample (e.g. class or school). This approach assumes that performance differences will reveal differences in difficulty between forms. Naturally, such designs rely heavily on the assumption that groups will be randomly equivalent.

The single group design requires that one group take both forms of the test to be equated. A common feature of such designs involves counterbalancing to adjust for any error introduced by order of administration. The advantage of such a design is that the

equivalence of groups is not an issue. However, a serious practical limitation is that it is potentially time consuming and expensive for examinees to take both complete forms.

Unlike single group designs where each form is taken by all examinees, the common-item nonequivalent group (CINEG) approach relies on all examinees taking only a representative subset of items common to both forms. That is, a set of items (often termed 'anchor' or 'link' items) are embedded in each form. All other items on the form are unique. This approach is practically appealing as it maintains the security of non-common items between groups and each group is only required to take one form. Examinee performance on the common items provides the basis for the equating adjustment. In this investigation a CINEG design will be employed.

Pre-Equating

Although there is a growing body of research on IRT based measurement models and equating procedures, there is a paucity of research on the efficacy of *pre-equating* with IRT. Pre-equating, as described earlier, simply describes the condition where items estimates are obtained prior to a live or operational administration. By so doing, number correct or true score to ability estimates can be said to be 'known' before a test is administered. To be fair, *pre-calibration* might be a more accurate term to describe this process. Nevertheless, the term pre-equating has found its place in the lexicon of applied psychometrics and will be used throughout this study in the manner described.

One of the first accounts of pre-equating methods is offered by Lord (1980). Lord describes a procedure for building a pool of pre-equated items such that "each new test form is equated to previous forms before it is administered" (p. 205). While some

of the advantages and limitations of this approach are considered in this brief treatment of pre-equating, no specific investigation into the efficacy of pre-equating is presented.

A review of the literature reveals only a few applied studies directly related to exploring the effectiveness of pre-equating. Livingston (1985) examined the effectiveness of pre-equating a large scale state assessment. The results were promising, showing relatively small prediction errors for most examinees. However, results were less accurate for examinees at lower ability levels, suggesting a need for further work. Moreover, Livingston's study was based on regression and equipercentile methods, which is not consistent with most contemporary approaches.

Similarly, Beard et al. (1984) examined the effectiveness of pre-equating domain subtests. The authors report promising results using Rasch model based equating methods. However, a few limitations of this study limit the generalizability of the findings. First, the calibrations were based on a relatively small sample of approximately 250 respondents. Second, there was overlap between the original pre-equating sample and the comparative sample. Therefore, the usefulness of this study for determining the effectiveness of pre-equating based on a different group of examinees is limited.

A more thorough examination of pre-equating was conducted by Stocking and Eignor (1986). This study utilized methods that have direct bearing on the present investigation. Hypothesizing that errors from pre-equating can be caused by 1) differences in the ability of respondents 2) multidimensionality of the data and 3) a combination of ability differences and multidimensionality, the authors conducted a simulation study to determine the effects of each condition. Samples were simulated to conduct the pre-equating and compared to the "true" values from a previously

administered form of the Scholastic Aptitude Test (SAT). In brief, the authors found that small differences between the ability distribution of the simulated samples and the true distribution were negligible, "these equatings differ by about what one would expect on the basis of estimation errors alone" (p. 23). However, large differences in ability led to errors beyond those associated with estimation and could distort the predicted raw to ability score relationship. Such differences led to scale score differences of up to five points.

The results from the multidimensional simulation were more dramatic. Stocking and Eignor found that when multidimensionality was introduced in the simulated data the *b* parameters were greatly overestimated. This led to equating errors of up to 30 scale score points. Similar results, but not quite as extreme, were found when both ability differences and multidimensionality were introduced. Clearly, these findings suggest that additional research into the efficacy of pre-equating is needed.

Measurement Equivalence

When respondents with the same scores on the latent trait have the same expected raw or true score at the item level, measurement equivalence is said to exist (Drasgow & Kanfer, 1985). According to Raju et al., "without measurement equivalence, it is difficult to interpret observed mean score differences meaningfully" (2002, p. 517). Therefore, it is critical to determine if a test functions the same for students in a field test administration versus an operational event. It is hypothesized that violations of the assumption of measurement equivalence will lead to instability in pre-equating.

There are two primary methods for analyzing measurement equivalence. First, various techniques based on Confirmatory Factor Analysis (CFA) can be applied.

Second, IRT based methods can be employed. Raju et al. demonstrated that there is "a high degree of congruence between the two methods" (2002, p.527). Importantly, CFA approaches are linear and IRT assumes a non-linear relationship between examinees and the latent trait. For this reason, an IRT based technique for assessing measurement equivalence was selected for this investigation.

Within an IRT framework one method of evaluating measurement equivalence is to investigate Differential Item Functioning (DIF). There are several IRT approaches to DIF such as Lord's chi-square (1980), Thissen, Steinberg, and Wainer's (1988) likelihood ratio tests, area measures proposed by Kim & Cohen (1991) and Raju (1990), and Raju, van der Linden, and Fleer's differential functioning of items and tests (DFIT) framework (1995). The DFIT method will be used in this study.

Among the appealing characteristics of the DFIT framework is that two kinds of DIF indices are produced: noncompensatory differential item functioning (NCDIF) and compensatory differential item functioning (CDIF). NCDIF is similar to other IRT based DIF indices, because it is based on the assumption that all other items on the test, except the one under investigation, do not exhibit DIF (Raju et al, 1995) NCDIF values are always non-negative, accordingly it does not distinguish between DIF that favors the focal or reference group. DFIT provides another index, CDIF, that does take into account bias from other items on the test, which may be more practical. CDIF provides an indication of the *direction* of DIF indicating which group is advantaged by the difference. Moreover, the sum of CDIF values equal to the DTF index, enabling practitioners to better evaluate the contribution of each item to DTF and the effect of removing items.

As described in Raju et al. (1995) the DFIT technique compares the probability of success (*P*) for examinee (*s)* with ability (θ) on item (*i)* when considered as a member of the focal group (*F*) to the probability of success when considered a member of the reference group (*R*). If an item functions differently $P_{iF}(\theta_s)$ and $P_{iR}(\theta_s)$ will be different for some examinees.

The NCDIF index for an item (*i*) can be represented as follows:

$$NCDIF_i = E_F[P_{iF}(\theta_s) - P_{iR}(\theta_s)]^2 = E_F(d_i^2) = \sigma_{d_i}^2 + \mu_{d_i}^2 \tag{2.8}$$

where the expectation (*E*) is taken over the focal group with respect to the squared difference of $P_{iF}(\theta_s)$ and $P_{iR}(\theta_s)$, as defined above. The d_i indicates the difference in true scores on item *i* for the same examinee, regarded as a member of the focal group and then as a member of the reference group. The standard deviation and mean are indicated by σ and μ, respectively.

At the test or subscale level, the difference in true scores (*D*) for an examinee can be expressed as the sum of d_i for all items on the test as indicated below:

$$D = \sum_{i=1}^{n} d_i \tag{2.9}$$

Therefore, DTF can be defined as:

$$DTF = E_F(D^2) = \sigma_D^2 + \mu_D^2 \tag{2.10}$$

Raju et al (1995) further define the CDIF index at the item level as:

$$CDIF_i = E_F(Dd_i) = Cov(D, d_i) + \mu_D \mu_{d_i} \tag{2.11}$$

where Cov indicates the covariance. Finally, as indicated above, DTF can be computed as the sum of CDIF over all test items as shown:

$$DTF = \sum_{i=1}^{k} CDIF_i \tag{2.12}$$

An important element of the DFIT framework is the determination of appropriate criteria to evaluate the statistical significance of DIF and DTF values. A χ^2 based significance test for NCDIF and DTF is detailed in Raju (1995). Moreover, Fleer proposed a cut-off criterion of .006 for both indices to address false positives that can result from oversensitivity of the χ^2 test to large sample sizes (1993). Fleer's criterion was based on a Monte Carlo study in which the .006 criterion was shown to falsely identify about one percent non DIF items as exhibiting significant DIF. Additionally, Flowers, Oshima, and Raju (1999) and Bolt (2002) proposed procedures to identify a cutoff by examining error and power for generated cutoff values. This work also included extensions to polytomous models. Although these procedures for deriving a cutoff value generally worked well, they are not trivial to produce and validate, which, until recently, posed a continuing challenge for many practitioners using the DFIT framework.

This issue seems to be remedied due to the recent work of Oshima, Raju, and Nanda (2006). Building on the earlier studies described above, Oshima et al. have proposed an item parameter replication (IPR) method for assessing the statistical

significance of NCDIF. Oshima et al. explain:

> In this new method, a cutoff score for each item is determined by obtaining a (1 - α) percentile rank score from a frequency distribution of NCDIF values under no DIF condition by generating a large number of item parameters based on the item parameter estimates and their variance-covariance structures from a computer program such as BIILOG-MG3. This cutoff for each item can be used as the basis for determining whether a given NCDIF index is significantly different from zero (p.2).

Evidence obtained from a Monte Carlo study is presented to validate these findings. In this investigation the IPR methods proposed by Oshima et al. and operationalized by Raju's DFITD7 Fortran program (2005) are used to evaluate the statistical significance of DIF and DTF.

CHAPTER 3

METHODS

Instrument

This investigation will use data from the Georgia Criterion Referenced Competency Test (CRCT). The CRCT is a state-wide, criterion-referenced assessment that measures mastery of the Georgia Quality Core Curriculum (QCC) for elementary and middle grades students in Georgia's public schools. Students in grades one through eight are administered the reading, English/Language Arts (ELA), and mathematics CRCT. Additionally, students in grades three through eight take the science and social studies CRCT. The test is typically administered in mid-April or early May to approximately 120,000 students in each grade.

Data from CRCT administrations in ELA, mathematics, and social studies for grades three, five, and seven will be used in this study. These grades and content areas were selected to investigate effects for three different constructs across three distinct age groups. Also, examination of these groups is more practically feasible, because all the link items for these tests were embedded in the spring 2002 operational test and the fall 2003 field test. This is not true of all grade and content combinations.

Student performance on the CRCT is reported as both a scale score and a performance level classification. Scale scores range from approximately 150 to 450. A score below 300 is termed Performance Level One (PL1), and indicates that a student did

not meet expectations. A score of 300 to 349 is termed Performance Level Two (PL2), which classifies a student as meeting expectations. Scores of 350 and above are termed Performance Level Three (PL3) and designate that a student has exceeded expectations. Students in PL2 or higher are regarded as performing on grade level.

Procedures

Phase I

First, item calibrations were obtained for the spring 2002 operational CRCT. Items were calibrated based on the Rasch model using Winsteps 3.49 (Linacre & Wright, 2004). These item calibrations provided "baseline" information for the link items later embedded in the fall 2003 field test and spring 2004 operational test. Essentially, the 2002 calibrations were used to establish the scale to which future calibrations will be equated.

In December of 2003 an off-grade CRCT field test was conducted in grades two through nine. In this field test students one grade above the target student population were administered the examinations (e.g. students in grade four took CRCT items intended for grade three students etc.). This design was selected to insure that all students taking the exam had at least one full year of instruction on the material covered in the assessments. The field test involved approximately 2,000 students in each grade and content area. To facilitate the common item equating design, items selected from the spring 2002 assessment were embedded on the fall 2003 field test as shown in Table 1. (Note: There was not an administration in fall 2002 or spring 2003 to use for comparison.)

Table 1

CRCT Content Test	Number of Forms	Items Per Form	Link Items Per Form
ELA	3	60	11-12
Mathematics	3	70	12-14
Social Studies	3	70	13-14

Description of Forms and Item Information for 2003 Field-Test

The next step in the process was to calibrate the field test items and equate them to the scale of the spring 2002 test. A common-item non-equivalent groups (CINEG) equating design was employed. The fixed *b* or anchor item method was used to scale the field test items to the reference test using Winsteps 3.49 (Linacre & Wright, 2004). It is important to note that three forms in each content area were field tested in fall 2003. On each form the same common link items from the reference test were embedded. Therefore, a total of nine forms with 600 items for each of three grades (1800 total items) were calibrated and equated as described above. However, once the full set of field test item calibrations are obtained, only one form for each grade and content area was created using these item parameters. These represent the forms that were used for spring 2004 operational testing.

To accomplish pre-equating of these forms, raw to scale score tables were created using the equated item parameters obtained for these forms. The scale scores for the CRCT were obtained by a simple linear transformation of θ where the scale score values of 300 and 350 correspond to the theta values for each cut-point. Given that θ_1 = cut-point for PL1/PL2 and θ_2 = cut-point for PL2/PL3, the CRCT scale score (SS) transformation is computed as:

$$SS = m(\theta) + b \tag{3.0}$$

Where

$$m = \frac{(350 - 300)}{(\theta_2 - \theta_1)}$$

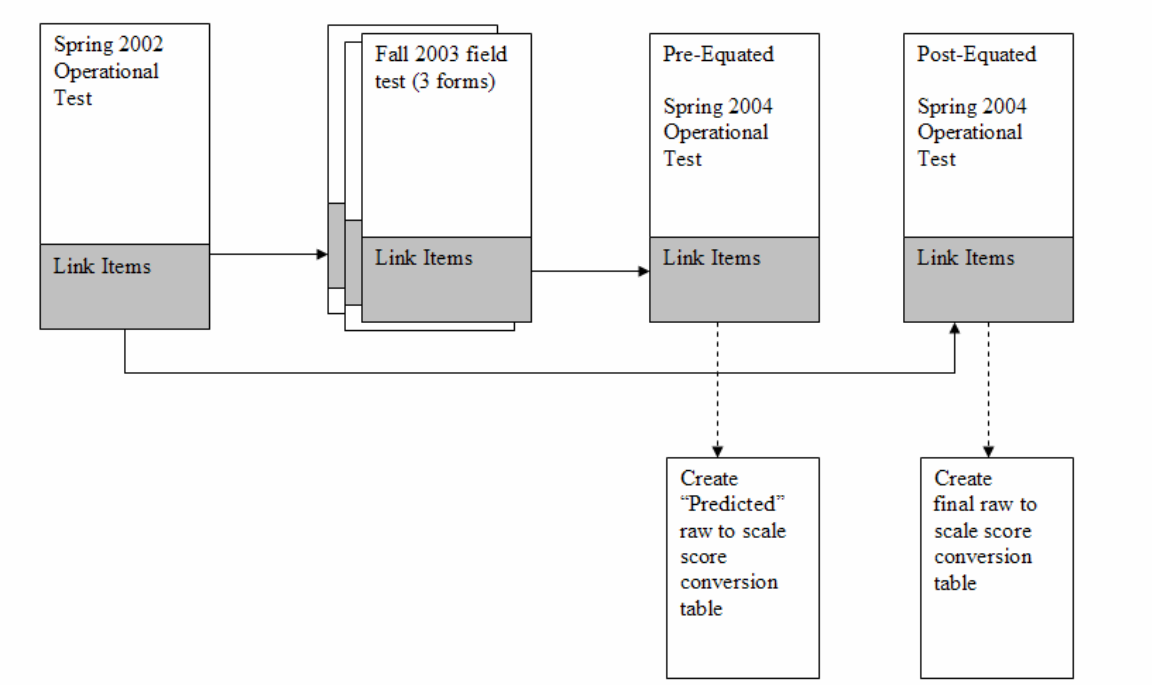
and

$$b = 350 - m(\theta_2)$$

The theta cutscores (θ_1 and θ_2) for the CRCT were established at a standard setting event in the summer of 2002. A modified-Angoff procedure was used to establish the standards. In brief, this process involved a group of panelists in each grade and content area who evaluated each item on the test with respect to the performance level descriptors (PLD) established by the Georgia Department of Education. The panelists were asked to consider what percent of students who just barely meet the state's expectations described by the PLDs would get the item correct. Similarly, they considered what percent of students who just barely exceeded expectations would get the item correct. In a series of individual and group ratings, the panelists essentially assigned each item a p-value or percent correct score. The p-values for each item were then summed for the all items on the test to obtain a raw or number correct score for each performance level. The theta value associated with this number correct score is regarded as the standard for meeting or exceeding expectations.

The next step in the process was to calibrate the items again following the full state-wide operational administration of the CRCT in the spring of 2004. Item parameters were obtained as described previously. Following this calibration, post-equating was conducted to evaluate the efficacy of the pre-equating. Post equating involved replicating the equating procedure described previously for each of the tests with the full population of state examinees. In this case the spring 2002 operational test remained the reference test and the spring 2004 test served as the focal assessment. Once the 2004 operational test was post-equated, the raw to scale score transformation tables were created as described previously. This design is described in Figure 1.

Figure 1



Overview of Study Design

To evaluate the efficacy of pre-equating with the field test data a number of analyses were conducted. First, the raw score cut-points from the pre-equated tables were compared to those obtained in post-equating. The post-equated tables are regarded as expressing the "true" relationship between the raw score and the scale scores, as these data are based on the entire population of examinees. Of primary interest is whether or not and/or to what extent the cut-points for PL2 and PL3 differed between the pre-equated tables and the post-equated tables. Obviously, if different cut-points are achieved with post-equating, students would have been misclassified by using the pre-equating information from the field test. The percent of students falling into each classification are compared for all grades and content areas investigated.

Second, differential item functioning (DIF) analyses are conducted to better understand where and why differences may occur between the field test and the operational test. The DFIT model implemented by DFITD7 was used (Raju, 2005). It was expected that the tests where DTF is significant and/or where a greater number or percent of items are flagged for DIF would lead to greater inconsistencies between pre-equating and post-equating. The DIF results were scrutinized for patterns by grade, content area, and domain. The goal was to provide information concerning what types of items are most likely to promote or degrade stability when pre-equating.

Because the parameters for the link items are held constant for all calibrations in Phase I, additional analyses were conducted to examine DIF in the link set. Parameters for the reference group were obtained by calibrating just the link items from the 2003 field test for all three forms simultaneously. A random sample of 1000 cases from the 2004 data were used for the focal group. These link sets were equated using the TCC

method operationalized by the computer program Equate 2.1 (Baker, 1993). Again, the DFIT model was used to evaluate DIF for these link items.

Finally, it was of interest to evaluate differences in item difficulty for the 2003 field test population compared to the 2004 operational population. It is not useful to evaluate the item calibrations for this purpose, as the *b* parameters are equated. Therefore, p-values for each item were produced for the off-grade 2003 field test and for the full population of examinees for the 2004 operational administration. Differences in p-values for all items between these two events were evaluated for direction and pattern.

Phase II

In a second phase of the study the item calibration and scoring procedures described in phase I were replicated with the one, two, and three parameter logistic IRT model for one content area. A few procedural differences were required for this phase. First, the program Bilog MG 3.11 was used to generate item calibrations (Mislevy & Bock, 1997). All examinees were included in the 2004 Bilog calibrations by using the command 'TAKE = *n*', where *n* indicates the number of examinees in the population. For the three parameter calibrations, the command 'COMMON' was used to set all *c* parameters to a common value. Moreover, the TCC equating method was employed for equating using eleven ability scale points.

Generation of score tables presented a unique challenge in this phase because there is not a one to one relationship between number correct scores and ability estimates in the two and three parameter model. For this reason a test characteristic curve was generated for each model to express the relationship between examinee true score and the

theta estimate for each true score. The program PIE was used to generate the true score to theta tables (Hanson & Zeng, 1995).

CHAPTER 4

RESULTS

Phase I

Pre-equated and post equated item parameters were obtained for each of the nine grade and content area tests used in this investigation. To evaluate the overall difference between the parameters, the Root Mean Square Difference (RMSE) was obtained for each set of values. RMSE is calculated as the square-root of the sum of the squared differences, divided by the total number of items compared (i.e. test length). All item parameter tables are presented in Appendix A and a summary of the RMSE differences is shown in Table 2 below.

Table 2

	Content Area		
Grade	ELA	Math	Social Studies
3	0.2297	0.3995	0.3036
5	0.3297	0.3167	0.4872
7	0.3401	0.3085	0.2596

Summary of RMSE of Item Parameter Estimates

Interestingly, there is no distinct or uniform patter to the error. It seems to increase in ELA from grades three to seven, whereas it decreases in math across the same grades. The greatest RMSE between pre-equated and post-equated item difficulty estimates was found in grade five social studies, which exceeds that of grade three and grade seven.

The direction of the differences in individual item parameter estimates (Appendix A) is also without a strong pattern. There are nearly an equal number of positive differences as negative differences. That is, some items were more difficult when calibrated based on an off-grade field test sample (pre-equated), compared to the calibrations obtained on-grade during the high-stakes operational administration (post-equated). Other items were less difficult for the field test sample.

A key part of this investigation involved creating the score tables for the pre-equated and post-equated models. By comparing the score tables, one can observe if there would have been a difference in assessment results and, critically, performance level determinations for students, based on equating methods. Score tables for all tests are shown in Appendix B and graphs of these tables are in Appendix C. The tables shown in Appendix B are shaded at the cut scores for "Meets Expectations" (300) and "Exceeds Expectations" (350) to more easily discern instances where performance level classifications would have been different depending on which equating approach was used. RMSE was calculated for the theta differences between each model to get an overall measure of difference. These values are shown in Table 3 below.

Table 3

	Content Area		
Grade	ELA	Math	Social Studies
3	0.0496	0.0517	0.1156
5	0.0455	0.0300	0.0231
7	0.0868	0.0199	0.0236

Summary of RMSE of Theta Estimates

Again, there appears to be no strong pattern to the differences in RMSE across grades and content areas. However, with the exception of seventh grade ELA, RMSE estimates are higher across content areas in grade three compared to grades, five and seven. Overall, the differences are remarkably small and similar.

A practically significant aspect of this study involved examining the performance level classification of students based on each equating approach. That is, how accurate is the pre-equating approach in determining whether a student will meet expectations or exceed expectations? Table 4 shows the percent of students classified in each performance level for each model. Naturally, in all instances where the cutscore is the same for pre-equating and post-equating the classification error is zero. There were three grade/ content areas where different passing cutscores were obtained. In grade three social studies and grade seven English/ language arts 1.71% and 2.17% of students were erroneously classified as passing (Type I error), whereas in grade seven social studies 3.03% of students that would have achieved PL2 based on post-equating failed using the pre-equating results (Type II error). Regarding 'exceeds expectations' performance, the pre-equating approach led to one Type I error in grade three social studies of 3.04% and two Type II errors in grade five math and grade seven social studies where, respectively, 2.75% and 2.13% of students were not classified in PL3 on the assessment. Although these errors are relatively small, it is important to note that about 115,000 students take this assessment in each grade. Therefore, a classification error of even 3% will yield about 3,500 erroneous performance level determinations.

Table 4

		Pre Equated		Post Equated		Classification Error	
		Meets*	**Exceeds**	**Meets***	**Exceeds**	**Meets**	**Exceeds**
Grade 3	**ELA**	85.28	19.85	85.28	19.85	0.00	0.00
	Math	85.17	20.02	85.17	20.02	0.00	0.00
	Social Studies	83.30	17.29	81.59	14.25	**1.71**	**3.04**
Grade 5	**ELA**	80.23	15.39	80.23	15.39	0.00	0.00
	Math	79.13	12.88	79.13	15.63	0.00	**-2.75**
	Social Studies	83.36	11.39	83.36	11.39	0.00	0.00
Grade 7	**ELA**	78.92	17.75	76.75	17.75	**2.17**	0.00
	Math	71.99	15.61	71.99	15.61	0.00	0.00
	Social Studies	70.55	14.98	73.58	17.11	**-3.03**	**-2.13**

* Includes Exceeds

Summary of Classification Accuracy

The next step in this investigation involved generating DIF and DTF indices for all items and tests. Before analyzing DIF and DTF, preliminary analyses were conducted to gauge the effectiveness and sensitively of DFITD7. Using the 2004 operational data file, randomly selected groups of 5000 and 1000 examinees were selected in third grade ELA. Items were calibrated using Winsteps 3.49 for each group. Item parameters were equated using the TCC method with Equate 2.1.

The DFITD7 output for the group of 5000 examinees revealed no groups exhibited significant DIF at the .001 level, one item was flagged at the .01 level, and three items were flagged at the .05 level. For the group of 1000 examinees five items were flagged at the .05 level and three were identified as exhibiting significant DIF at the .001 level. While the results were similar, the sample size of 1000 was slightly more sensitive. Therefore a sample size of 1000 was selected for this study. The full DIF results are presented in Appendix D.

For the DIF analyses the reference group was based on the parameters obtained from the 2003 field test and the focal group was based on parameters and variances from

1000 randomly selected cases from the 2004 operational test. The parameters based on the random sample (n=1000) were found to be very strongly correlated to those from the full population (n≈115,000) as shown in Table 5. All randomly generated item parameters and those from the full population can be found in Appendix E.

Table 5

Test	Correlation
Grade 3 ELA	0.99519
Grade 3 Math	0.99639
Grade 3 Social Studies	0.99556
Grade 5 ELA	0.99678
Grade 5 Math	0.99525
Grade 5 Social Studies	0.99760
Grade 7 ELA	0.99795
Grade 7 Math	0.99662
Grade 7 Social Studies	0.99353

Correlation of Item Difficulty Estimates From Random Sample to Population Parameters

Given the congruence in score tables based on pre-equated and post-equated conditions, the DIF findings were somewhat surprising. Generally, a fifth or more of the items on each form were flagged for DIF between pre-equating and post-equating. The proportion of DIF items ranges from 18% in seventh grade ELA to 42% in third grade math. Again, no discernable pattern emerges regarding the number or percent of DIF items. Moreover, the areas with a greater percentage of DIF items do not appear to be related to the size of RMSE item parameter differences or RMSE theta differences. For example, grade three social studies had the highest RMSE value for theta differences, but has one of the lower percentages of DIF items (32%) compared to the other forms. Table 6 shows the number of items flagged for DIF at the .001 level of significance. DIF and DTF information for each item and test is provided in Appendix F.

Table 6

	Content Area		
Grade	ELA	Math	Social Studies
3	17 (34%)	25 (42%)	19 (32%)
5	10 (20%)	24 (40%)	19 (32%)
7	9 (18%)	21 (35%)	23 (38%)

Frequency/ Percent of Items Demonstrating DIF at .001

It may be more useful, however, to examine DTF for each grade and content area as an indication of measurement equivalence. The DTF indices are shown in Table 7 below. Six of the nine forms had a highly significant DTF value ($p<.001$), one form had a significance level that was slightly less pronounced ($p = .007$), while the DTF for two forms was not statistically significant. The two non-significant forms were fifth grade social studies and seventh grade math. Interestingly, these were also the content areas with the most congruence with respect to the score tables. That is, the RMSE of the theta differences at each true score value were lowest for the forms where DTF was not significant. This pattern appears to hold throughout. Higher DTF values are associated with higher RMSE theta differences. For example, third grade social studies had a particularly high DTF (2.9996) and also had an usually high value for RMSE theta difference, compared to the other forms.

Table 7

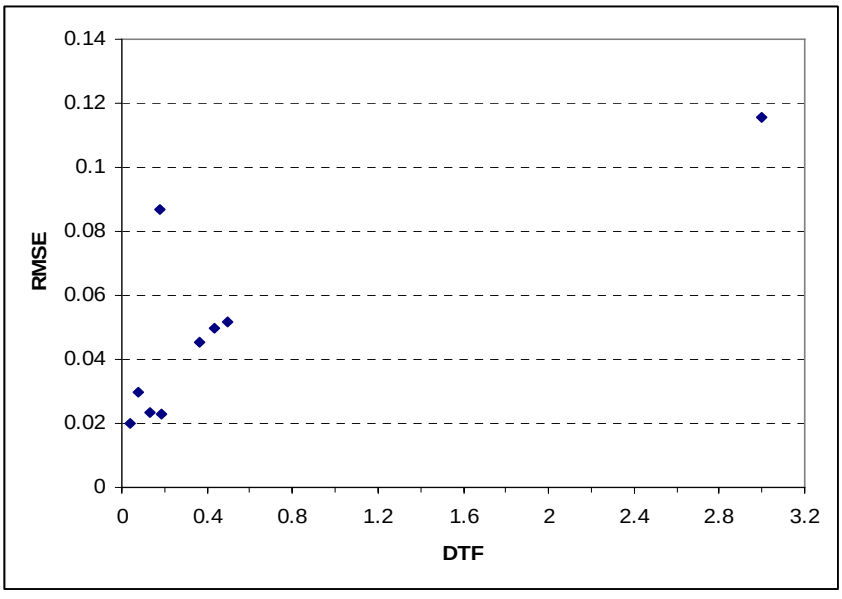
	Content Area		
Grade	ELA	Math	Social Studies
3	.4305**	.4928**	2.9996**
5	.3637**	.0813*	.1891
7	.1755**	.0406	.1336**

** p < .001, * p < .01

Summary of DTF

The relationship between DTF and RMSE of theta differences is shown graphically in Figure 2 below. This plot reveals a positive, linear relationship between DTF and RMSE for the nine grade content areas in this investigation (r =.797). These findings suggest that the extent to which two tests depart from measurement equivalence, will determine the extent to which the score tables for these tests will be incongruous.

Figure 2



Relationship between DTF and RMSE of Theta Difference

To further understand the factors that may contribute to a lack of measurement equivalence, DIF results were broken down by subtest or domain. Domains serve to group items by related skills and curricular standards. Tables 8, 9, and 10 show the number and percent of DIF items at the .001 level by content area, grade, and domain. These findings are also shown graphically in Figures 3, 4, and 5.

Table 8

Grade\Content	Domain	No DIF	DIF	% DIF
Grade 3 ELA	Grammar and Mechanics	14	5	26.3
	Paragraph Content and Organization	8	2	20.0
	Research Process/Source Materials	3	4	57.1
	Sentence Construction and Revision	8	6	42.9
Grade 3 ELA Total		33	17	34.0
Grade 5 ELA	Grammar and Mechanics	14	1	6.7
	Paragraph Content and Organization	7	3	30.0
	Research Process/Source Materials	8	2	20.0
	Sentence Construction and Revision	11	4	26.7
Grade 5 ELA Total		40	10	20.0
Grade 7 ELA	Grammar and Mechanics	13	1	7.1
	Paragraph Content and Organization	9	3	25.0
	Research Process/Source Materials	9	1	10.0
	Sentence Construction and Revision	10	4	28.6
Grade 7 ELA Total		41	9	18.0

Number and Percent of ELA DIF Items Grade and Domain

Table 9

Grade\Content	Domain	No DIF	DIF	% DIF
Grade 3 Math	Computation and Estimation	5	7	58.3
	Geometry and Measurement	7	6	46.2
	Number Sense and Numeration	10	4	28.6
	Patterns and Relationships/Algebra	5	1	16.7
	Problem Solving	6	3	33.3
	Statistics and Probability	2	4	66.7
Grade 3 Math Total		35	25	41.7
Grade 5 Math	Computation and Estimation	7	5	41.7
	Geometry and Measurement	7	3	30.0
	Number Sense and Numeration	5	8	61.5
	Patterns and Relationships/Algebra	5	2	28.6
	Problem Solving	8	4	33.3
	Statistics and Probability	5	1	16.7
Grade 5 Math Total		37	23	38.3
Grade 7 Math	Computation and Estimation	5	3	37.5
	Geometry and Measurement	9	3	25.0
	Number Sense and Numeration	5	4	44.4
	Patterns and Relationships/Algebra	9	4	30.8
	Problem Solving	8	3	27.3
	Statistics and Probability	3	4	57.1
Grade 7 Math Total		39	21	35.0

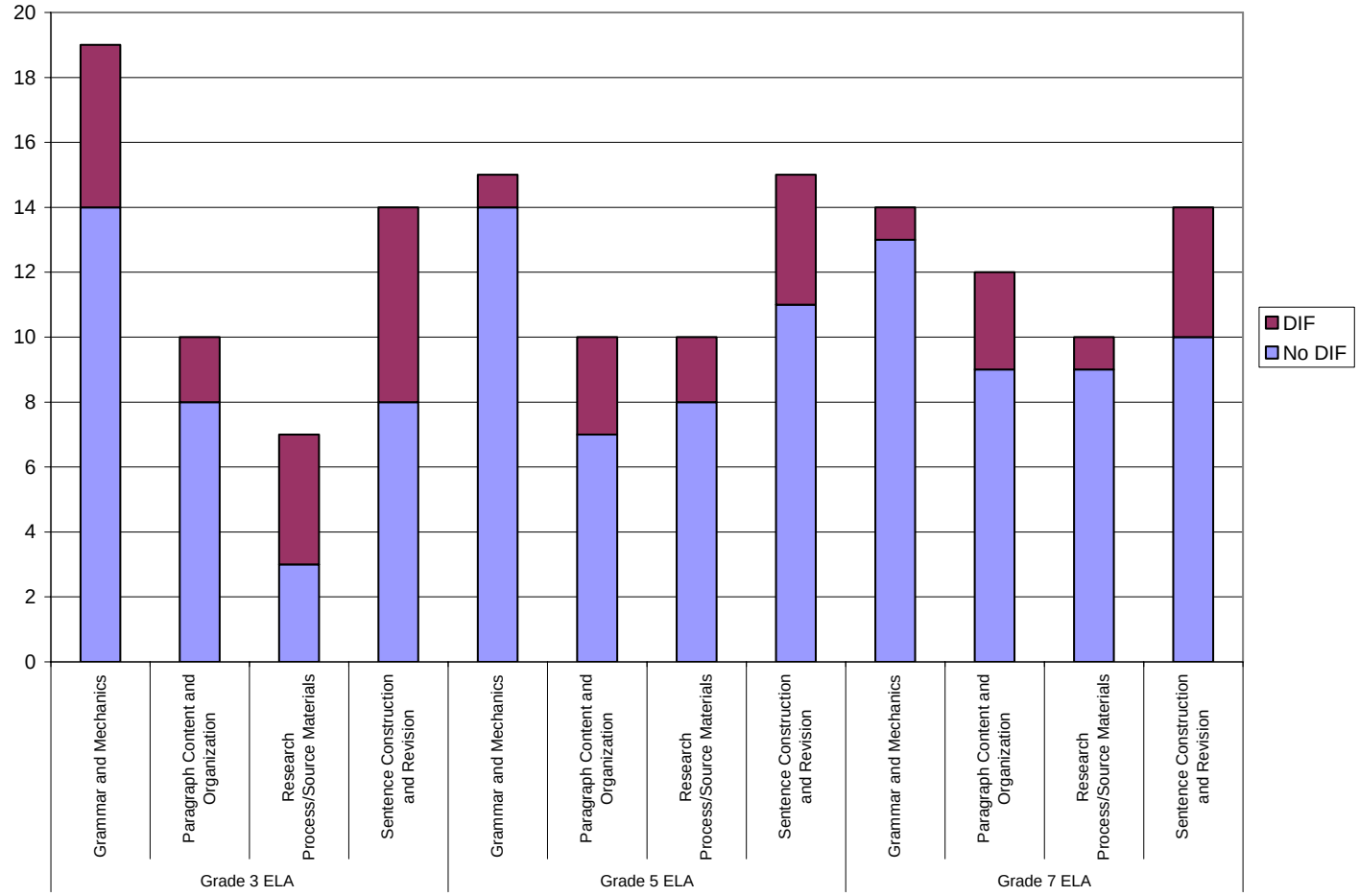
Number and Percent of Mathematics DIF Items Grade and Domain

Table 10

Grade\Content	Domain	No DIF	DIF	% DIF
Grade 3 Social Studies	Civics	6	5	45.5
	Core Skills Stand Alone	3	3	50.0
	Economics	3	3	50.0
	Geography	17	7	29.2
	History	12	1	7.7
Grade 3 Social Studies Total		41	19	31.7
Grade 5 Social Studies	Civics	6	3	33.3
	Core Skills Stand Alone	6		0.0
	Economics	5	2	28.6
	Geography	9	3	25.0
	History	15	11	42.3
Grade 5 Social Studies Total		41	19	31.7
Grade 7 Social Studies	Civics	6	5	45.5
	Core Skills Stand Alone	3	3	50.0
	Economics	9	1	10.0
	Geography	10	7	41.2
	History	9	7	43.8
Grade 7 Social Studies Total		37	23	38.3

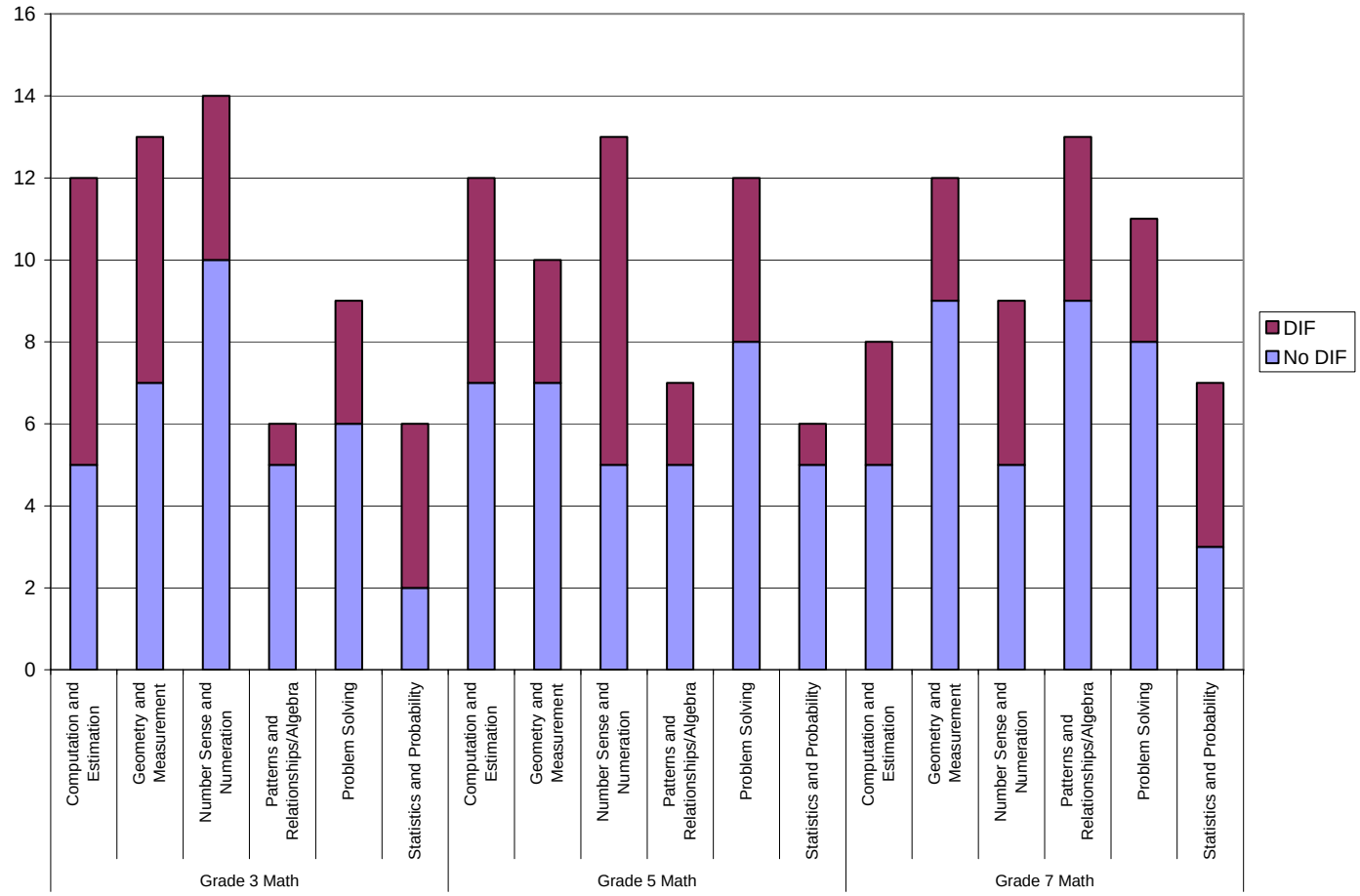
Number and Percent of Social Studies DIF Items Grade and Domain

Figure 3



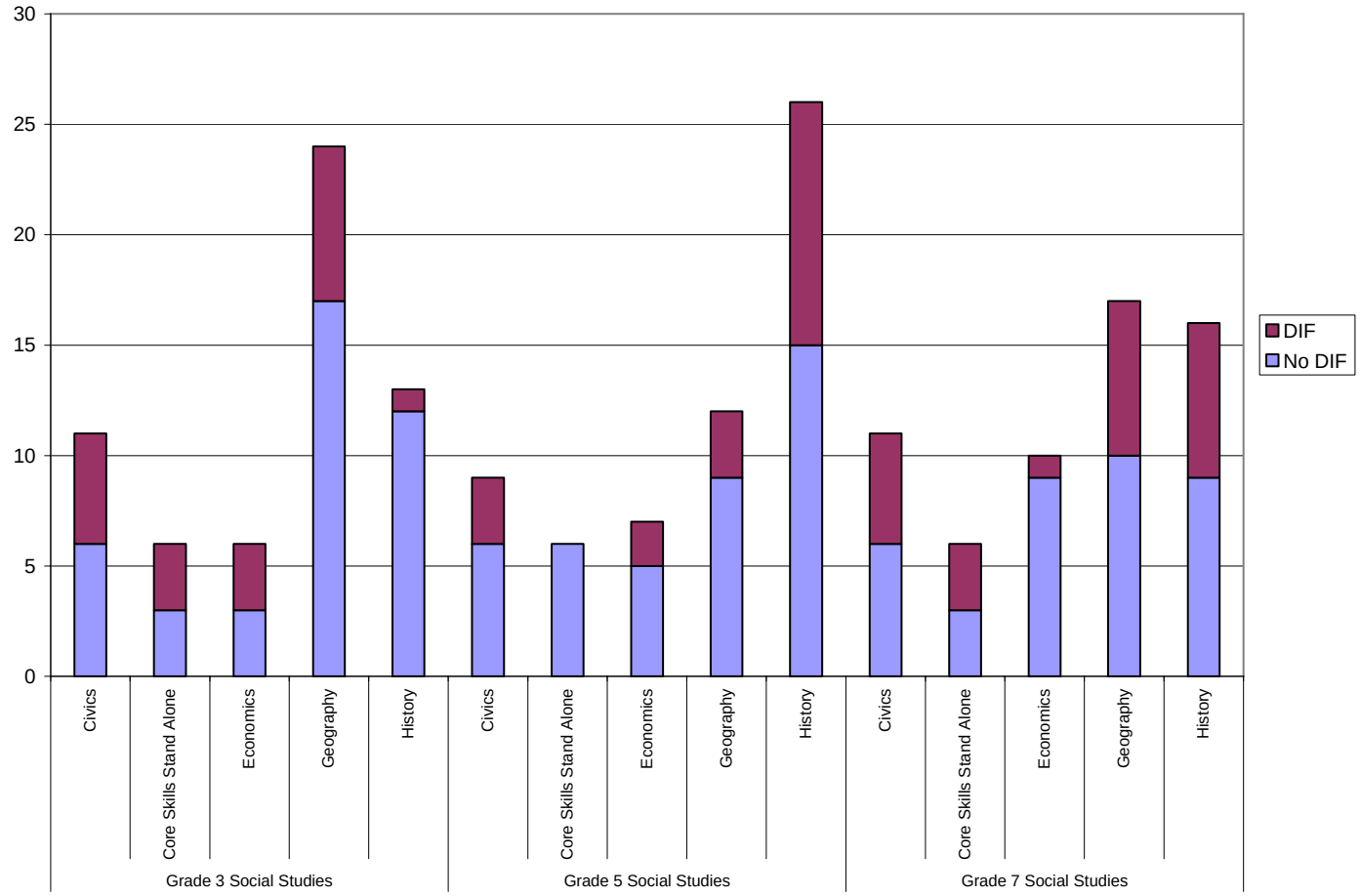
Graph of ELA DIF Items by Domain

Figure 4



Graph of Mathematics DIF Items by Domain

Figure 5



Graph of Social Studies DIF Items by Domain

Perhaps no firm conclusions can be reached from the domain analysis, but with some exception a few patterns do emerge. Generally, the fewest DIF items were found in ELA, but there seemed to be no pattern to which domains showed the most DIF within ELA across grades. In mathematics, patterns and relationships and problem solving generally contained the fewest DIF items. On the other hand, more than half of the items in statistics and probability in grades three and seven exhibited DIF, which exceeded DIF found from the other domains. In social studies, the domains generally contributed a comparable number of DIF items, with the exception of history in grade three, core skills in grade five, and economics in grade seven. In each case, these domains showed relatively few DIF items compared to the other domains.

The next set of analyses for Phase I involved investigating DIF for only the set of link items on each test. Item calibrations for the reference group were obtained by an unrestricted calibration of the link set for each of the nine tests across all three forms from the 2003 field test. The focal group was based on a sample of 1000 responses on the link set from the 2004 operational file. Table 11 shows the number and percent of DIF items flagged at the .001 level using DFITD7. Table 12 summarizes the DTF indices for each content area based on the set of link items. The full results of all link set DIF analyses are provided in Appendix G.

Table 11

	Content Area		
Grade	ELA	Math	Social Studies
3	1 (8%)	2 (17%)	3 (21%)
5	1 (9%)	4 (29%)	3 (21%)
7	2 (17%)	2 (14%)	6 (46%)

Frequency and Percent of DIF Items at .001 for Link Set

Table 12

	Content Area		
Grade	ELA	Math	Social Studies
3	0.00175	0.00461	0.00124
5	0.00178	0.00094	0.00349
7	0.01157	0.00501	0.01697

Summary of DTF for Link Items

The results of these analyses show that a larger number of DIF items are generally found in the link sets for mathematics and social studies, compared to ELA. Additionally, there is a clear pattern of increasing departure from measurement equivalence as the grade level of the tests increase. Notwithstanding, DTF indices were not significant for any of the nine link sets.

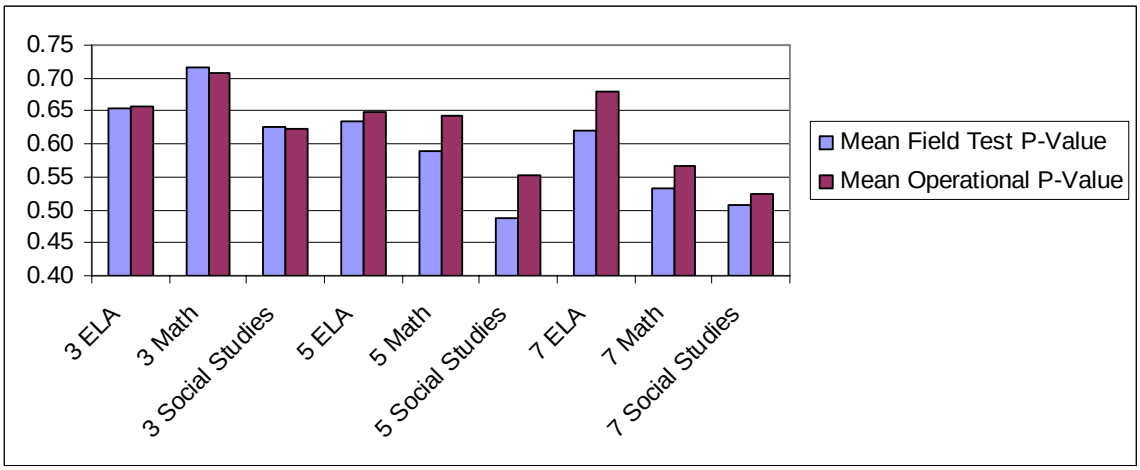
Lastly, p-values were produced for all items on each test to evaluate differences in difficulty between the 2003 field tests and the 2004 operational assessments. The results are summarized in Table 13 and Figure 6. A full table of all p-values for each assessment in each event are displayed in Appendix H. Moreover, this information is shown graphically as scatterplots in Appendix I.

Table 13

Grade/ Content	Mean Field Test P-Value	Mean Operational P-Value
3 ELA	0.6539	0.6581
3 Math	0.7148	0.7063
3 Social Studies	0.6259	0.6241
5 ELA	0.6337	0.6475
5 Math	0.5882	0.6422
5 Social Studies	0.4874	0.5530
7 ELA	0.6196	0.6790
7 Math	0.5327	0.5658
7 Social Studies	0.5075	0.5235

Summary of Mean Item P-Values

Figure 6



Graph of Mean Item P-Values

This summary information shows a clear difference in item difficulty for grade three compared to grades five and seven. Items are generally very close in difficulty in grade three with students most often getting a higher percentage correct on the field test compared to the operational test. However, this pattern reverses sharply in grades five and seven. In the upper grades, the field test items have substantially lower p-values than

do the operational items. It is most pronounced in grade five social studies where the mean p-value for the field test items was .4874 compared to .5530 for the operational test. The scatterplots in Appendix I, show this pattern with striking clarity. There is a random and fairly close scattering of difficulty differences in grade three. However, in grades five and seven, nearly every item has a higher p-value on the operational test compared to the field test. In fact, one item in grade five social studies had a field test p-value of .2102 and an operational p-value of .8390. This occurrence was highly exceptional, however, as p-value differences greater than .10 were otherwise rare.

Phase II

The grade three social studies test was selected for Phase II analyses, because it demonstrated the most pronounced difference between pre and post equating theta estimates. This test also had a DTF value more than six times larger than the next highest content area.

First, Bilog MG 3.11 was used to generate item parameter estimates for all 2003 field tests and for the 2004 operational test. The item parameter differences for the 1PL, 2PL, and 3PL are shown in Appendix J. A summary of these results for the *b* parameter are displayed in Table 14. Overall, the differences are negligible among all four models. There does seem to be a lower RMSD value for the 2PL and 3PL model, however, compared to the 1PL model. But, again, this difference is very modest.

Table 14

Model	RMSD of *b*
1PL/ Bilog	0.3763
2PL/ Bilog	0.3362
3PL/ Bilog	0.3250

Summary of Phase II b Parameter RMSD

Next the score tables for each model in Phase II were produced. The score tables are simply the Test Characteristic Curve (TCC) for each test. It was noted that the scale of the item difficulty estimates obtained from Bilog differed from those obtained from Winsteps. However, the correlation between the estimates was .9999, which indicates that, as expected, there is a linear relationship between the two.

Since it was necessary to adjust the theta cuts to reflect the differences in scales, the mean and sigma method was used to transform the scale produced by Winsteps (*X*) to that produced by Bilog (*Y*). Using the 2002 difficulty estimates produced by Winsteps and those produced for each of the 1PL, 2PL, and 3PL in Bilog, coefficients α and β were computed as follows:

$$\alpha = \frac{sd_y}{sd_x} \tag{4.8}$$

where: sd = standard deviation of the difficulty estimates of the items on form *Y* or *X* and

$$\beta = \bar{b}_y - \alpha \bar{b}_x . \tag{4.9}$$

where: $\bar{b}$ = mean of the difficulty estimates of the items on form *Y* or *X*

Then, the theta cuts were adjusted by:

$$\theta_y = \alpha\theta_x + \beta \qquad (4.10)$$

That is, a new theta cutscore for the 1PL, 2PL, and 3PL model was transformed to the scale of the Winsteps calibrations. In this manner, an equivalent standard was used in Phase I and Phase II of this investigation.

The resulting score tables are shown in Appendix K and graphs of these tables are presented in Appendix L. Table 15 summarizes the RMSD of the theta estimates for each model. These results suggest that the incongruence between pre-equating and post-equating is more pronounced using a more sophisticated model. In fact, the difference in the score table is more pronounced for the Phase II models compared to the Phase I model. The 3PL model has the largest RMSD of theta estimates followed by the 2PL, and 1PL, respectively.

Table 15

Model	RMSD of Theta
Rasch/ Winsteps	0.1156
1PL/ Bilog	0.1573
2PL/ Bilog	0.2006
3PL/ Bilog	0.3317

Summary of Phase II RMSD of Theta Estimates

DIF analyses were also run for each model in Phase II. Again, DIF was evaluated using the DFIT model implemented by DFITD7, this time for the 1PL, 2PL, and 3PL models for grade three social studies. The results are summarized in Table 16 and are presently fully in Appendix M. These results reveal that fewer items exhibited DIF for

the two and three parameter model (23) compared to the one parameter model (29). However in all cases, more items were flagged for DIF than found in Phase I, where 19 items demonstrated DIF for grade three social studies.

Table 16

Model	Freq (%) DIF
1PL	29 (48%)
2PL	23 (38%)
3PL	23 (38%)

Frequency/ Percent of Items Demonstrating DIF at .001

Similarly, DTF values were much higher for each model in Phase II compared to Phase I. These values are shown in Table 17 DTF was smallest for the two parameter model (4.2637) and largest for the one parameter model (6.7223). All DTF values were significant at the .001 level.

Table 17

Model	**DTF**
1PL	6.7223**
2PL	4.2637**
3PL	5.4431**

**P<.001

Differential Test Functioning for One, Two, and Three Parameter Models

CHAPTER 5

DISCUSSION

Measurement Equivalence

Overall, measurement equivalence for the majority of the grade and content areas was not established between tests administered to the off-grade field test sample in 2003 and the operational assessment administered in 2004. This is based on DTF values that were significant at the .01 level in seven of the nine grade and content areas evaluated. Only seventh grade mathematics and fifth grade social studies had a DTF index that was not statistically significant.

It is important to note, however, that DTF values were highest in grade three, and reduced in grades five and seven respectively, even while most remained significant. Similarly, RMSE of theta estimates were smaller for the score tables for grades five and seven, than for those of grade three. This suggests that lack of measurement equivalence is more pronounced for younger examinees than for older students.

One reason for this may be related to research about cognitive growth in young learners. Several studies indicate that cognitive growth does not approach 'stability' until about age nine or ten, which is typically a student's fourth grade year (Ausubel, Sullivan, & Ives, 1980; Rice, 1997). This suggests that students may be experiencing sporadic and/or unpredictable cognitive growth apart from formal schooling between the field test and the operational test. Stated another way, a young student may grow substantially with respect to a construct such as 'reading ability' or 'logical reasoning' between the

spring of their third grade year and late fall of their fourth grade year. Such sporadic and non-uniform cognitive growth in these areas is likely to influence how students perform on test items.

However, if cognitive growth due to chronological age difference was the only influence on performance, one would expect to see that all field test items were easier for these older students. This is clearly not the case. An examination of the p-values for the third grade tests shows that about an equal number of items had higher p-values on the field test compared to the operational test. In English/ language arts, for example, 22 items had higher p-values or were easier for the fourth graders who encountered them on the field test, compared to 23 items that had higher p-values for third graders on the operational test. Mathematics was nearly evenly split as well, with 31 items easier for field test examinees and 29 easier for the younger operational examinees. Interestingly, this pattern changes dramatically for students in grades five and seven. In these upper grades, items are consistently more difficult on the field test. In fifth grade social studies only 7 of 60 items has a higher p-value on the field test. Similarly, in seventh grade English/ language arts only 5 of 50 items are easier for the older students taking the field test. Indeed, this seems to contradict the hypothesis that older students perform better on the items due to cognitive growth. At the least, the data in this study suggest that this interpretation is not supported for students in fifth grade or higher. Overall, the most plausible explanation is that other factors, perhaps aside from or in concert with cognitive growth, impact performance between these two groups.

Recency of instruction may be one of these factors. Some research shows that student performance may regress as students are removed from instruction. For example,

Ceci (1991) argues that performance gains made during the school year may decline during summer vacations. One might attribute this to a 'recency effect', a term that is not unfamiliar in cognitive psychology to describe how memory fades with the passage of time (e.g. Matlin, 2002). Indeed, it seems plausible that a recency effect would have a greater impact on social studies as the data show. Social studies items are much more connected to a specific, discrete curriculum than are more general language arts or math items. Language arts and mathematics are also part of a more 'spiraled' curricula, meaning that each year students build on knowledge learned in previous years. This may include repeating much of the same content.

For example, in language arts students may learn to discern the main idea from a paragraph or reading passage. As the student advances in school, this skill continues to be taught, but the reading passages may increase in complexity. On the other hand, the social studies curriculum is more 'laddered' from grade to grade. Students may learn about American history one year and European history the following year. There is not a lot of overlap in content, consequently the student's knowledge may regress due to lack of recent instruction.

Still another factor that may impact student performance between administrations is motivation. Numerous studies have established a link between student effort and test performance (e.g., Weiner, 1990; Wolfe & Smith, 1995). A major factor that contributes to motivation is the perceived consequence of the test. Wolfe & Smith present research showing that students testing in consequential conditions outperformed students in a non-consequential conditions by an effect size of .26 (1995). They explain simply, "consequence influences motivation, and motivation influences performance" (p. 228).

Certainly, motivation is a likely contributor to performance differences found in this investigation between students taking the field test, compared to students taking the summative high stakes assessment. Indeed, it seems plausible that students taking the field test would not exert as much effort, since no stakes were associated with this test event and, in fact, no student level results were ever reported.

The data support this hypothesis, particularly for the older examinees. As noted, there were sharp differences in p-values between the low-stakes field test and the high stakes operational test, where students performed better. Indeed, the mean p-value of the items was higher for each of the six tests taken by students in grades five and seven. However, in grade three the mean p-value of items on the low stakes field test was higher, albeit very similar, for two out of three tests. This suggests that older students, who may be more aware of the consequences of the test, were influenced by motivation, whereas motivation effects for the younger students was much less of an issue. In any case, it may be that motivation can only be added to the list of potential explanations for group differences, but it certainly does not account for all variation in performance.

In summary, it seems likely that cognitive growth, recency, and motivation may have contributed to the lack of measurement equivalence between the field test and the operation test found in most conditions. However, it remains unknown how much any one of these factors is responsible for the differences. The explanation is almost certainly buried in the complex interaction of these and other factors that both advantage and disadvantage performance. Indeed, even when differences between groups at the item or test level are not pronounced, it may be because factors such as those described

canceled each other out. Additional research to disentangle the isolated effects of these factors would be very illuminating.

Score Table Comparisons and Classification Accuracy

One of the more interesting findings of this research is that the raw or true score to scale score tables were remarkably similar, even when DTF indices were significant. Moreover, results from pre-equated and post-equated tables frequently classified students in performance levels very consistently. From a technical or academic standpoint, it could be argued that it is most important to evaluate this finding by examining the differences in the theta estimates for each raw score value as shown in Tables 3 and 12. From an applied perspective, on the other hand, classification accuracy is they key concern. However, there are some additional factors to consider when evaluating classification that confound interpretation.

In large scale criterion-referenced achievement testing, a student is classified in a performance level (e.g. proficient, advanced etc.) if he or she achieves a number correct score that corresponds to an ability estimate (θ) that is equal to or exceeds the standard. This value is established at a standard setting event, as described in chapter three, and is then carried forward to each subsequent equated form. However, when another form is used, it is highly unlikely that the exact theta value established in standard setting will appear in the raw score to theta distribution. Indeed, with only 40 to 60 discrete points in this distribution for the CRCT, the 'standard theta' will almost always fall between two 'observed thetas'. There are a number of approaches for dealing with this. The approach selected is typically based more on policy than on psychometrics.

One approach is to simply create the linear transformation between the theta and the scale score using the same 'standard theta' in the equation (refer to equation 3.0). This was the approach used in this investigation. The advantage of this method is that it is arguably more precise because no rounding occurs. The disadvantage is that scores precisely at the cutoff will not appear in the distribution. It might present a policy concern if, for example, a student can achieve a score of 300 on test form A, but on test form B, the same raw score produces a 299 and the next highest score produces a 303. In such a case, an argument could be made that the student who took form B was held to a higher standard. Indeed, the 303 would certainly connect to a higher theta than the one established at standard setting.

This problem can be overcome by implementing a rounding rule. A conservative rule would hold that the closest theta in the distribution without exceeding the standard theta would replace the standard theta. That is, the closest observed theta is used in equation 3.0 to produce the scaled score. By so doing, a score of 300 would always appear in the distribution and no student would be held to a higher standard than was established at standard setting. Such an approach may be politically appealing, because it guards against false negative results, giving students the benefit of the doubt where high stakes decisions are involved. The disadvantage of such an approach is that the standard does 'move' insofar as it is rounded to the closest value. Such rounding can cause the theta cutoff to change by a non-trivial amount.

This information is presented in some detail because it has the potential to cloud understanding of the classification tables presented in this investigation. As noted earlier, the score tables produced from pre-equating and post-equating were found to be very

similar. This is evidenced in the small RMSE of theta differences (see Tables 3 and 12). Moreover, it can be seen graphically in the nearly coincident plots produced in Appendices C and H. However, even when theta values are very close for the same raw score, the performance level classification may differ depending on how close the 'observed theta' in the distribution is to the 'standard theta'.

Seventh grade ELA provide a good example of this point. Note in Appendix B that the theta value associated with 25 items correct on the pre-equated score table is .0275, which transforms to a scale score of exactly 300. For the post-equated table, the theta for 25 items correct is -.0151, which transforms to a scale score of 299. A student held to the post-equated table would have to earn a number correct score of 26, which maps to a 302, to pass the test. In this example, the 'standard theta' is .01 for PL2. The theta associated with a raw score of 25 on the post-equated table is actually closer to the standard theta (.02 compared to .06) but it does not yield the same classification. Importantly, the 'error' can work in the other direction as well. That is, score tables that may not be very close, may produce the same classification depending on the value of observed thetas in the distribution.

Notwithstanding these caveats it remains a practically important feature of this study to examine the consistency in performance level classification between pre-equating and post-equating. In three of the nine tests students were classified differently at the 'meets expectation' or PL2 cutoff score. The classification differences were no more than 3% and in two out of three cases favored the pre-equated score tables. That is, a student was more likely to pass based on the pre-equated score tables, than the post equated tables. Results were similar for the 'exceeds expectations' or PL3 cutoff score.

Again, three of nine tests produced classification error with the maximum error at 3%. Two of the same tests were affected: third grade social studies and seventh grade social studies. Seventh grade ELA and fifth grade mathematics differed at PL2 and PL3, respectively. However, with regard to PL3, the post-equated tables advantaged students in two out of three cases. While the classification errors seem small, as noted previously, statewide this can amount to about 3,500 erroneously classified students.

Overall, it does not seem likely that the classification results can reliably extrapolate to a more general finding about classification precision for pre-equated versus post-equated methods. First, as explained, other factors (e.g. scaling decisions) can contribute to the error of classification. Second, a clear or strong pattern was not evident in these findings. The classification errors were small, were not restricted to any one grade or content area, and were of both the Type I and Type II variety. The one exception is that four of the six errors occurred in social studies. This, along with additional evidence to be examined in more detail, may suggest reduced stability for this content area.

As indicated previously, the score tables were very similar with regard to the RMSE of the theta differences. In all instances, the difference in the estimated theta at the number correct cutscore between the pre-equated and post-equated tables was much less than the standard error of the theta estimate for the full population of examinees. The standard error consistently ranged from .2 to .3, whereas the highest theta difference at the cutscore was .11, which was found in third grade social studies. This evidence suggests, then, that the methods used to pre-equate the large scale criterion-referenced test studied in investigation do produce score tables that function equivalently to those

obtained through post-equating. Where the tables produced different classification decisions, it can be argued that this was more an artifact of the scaling transformation method than of equating error.

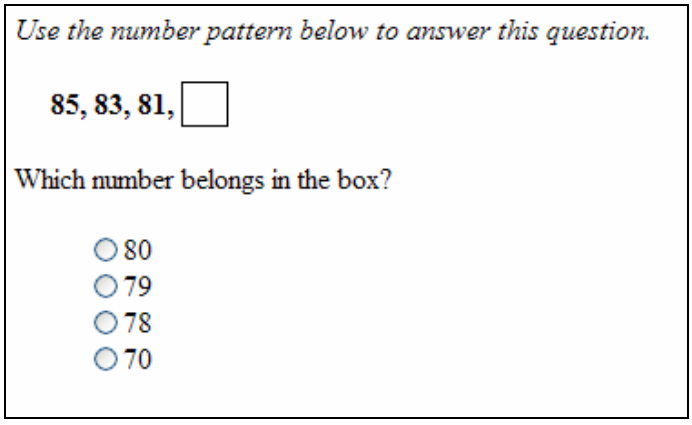
Items and Domains

The DIF findings provide some evidence that items measuring more 'generalized' aspects of the curriculum may function more consistently between field testing and operational testing. As noted previously, the fewest number of significant DIF items were found in ELA and for the most part the number reduced across all content areas as grade level advanced. Content areas in the earlier grades do tend to focus on more foundational aspects of the curriculum. Also, ELA may be the most generalized of the content areas explored in this study. Even for content areas that are less generalized, the domains that typically produced fewer DIF items are those that connect to the broadest aspects of the curriculum. For example, the domain of history in third grade refers mostly to the study of communities and the roles of citizens. However, in seventh grade, the history domain covers specific events in the history of the Middle East, Asia, and Africa. To be sure, the history domain is more discrete in grade seven than in grade three. It may not be surprising, then, that there were relatively few significant DIF items in this domain in grade three (8%), compared to a much larger proportion in grade seven (44%).

This trend is also observed in mathematics. The domain of patterns and relationships contained the fewest DIF items across all grades. This domain is associated with very general cognitive skills that may be less connected with discrete knowledge than other domains, such as statistics and probability. For example, a common item type

in the patterns and relationships domain requires the student to determine the next number in a sequence. An example of such an item for third grade mathematics is shown in Figure 7. It could be argued that the student's ability to correctly respond to this item is not as strongly connected to classroom instruction. Consequently, the item may function very consistency for students across grade levels and at different times of the academic year.

Figure 7



Example of a Third Grade Item in the 'Patterns and Relationships' Domain

Phase II Results

One important goal of this study was to determine if using a more 'complete' IRT model would produce more stability between results obtained in field testing compared to those from operational testing. The findings of this investigation show that this is not the case.

The procedures used for all nine tests with the Rasch model were repeated for third grade social studies using the 1PL, 2PL, and 3PL IRT model. In lieu of the 'fixed *b*' equating method in Winsteps, the TCC method was used and items were calibrated using Bilog MG. The findings show that the RMSE of the theta difference was smallest

for the 1PL and was very close to, but slightly higher than, the RMSE found in Phase I. RMSE theta increased with the 2PL and 3PL, respectively. Curiously, the DTF value was highest for the 1PL model, followed by the 3PL, and finally the 2PL in descending order of strength. All DTF values were statistically significant at the .001 level and all were much higher than those found in Phase I. This, along with the fact that only three values were computed, may explain why no pattern between DTF and RMSE of theta difference emerged in Phase II.

Moreover, the score tables showed the same or greater incongruence than the findings for third grade social studies in Phase I. At the PL2 cutoff score, differences of one to three true score points were observed. This difference was most pronounced for the 2PL model, followed by the 1PL and 3PL, respectively. In sum, the methods used in Phase II of this study indicate that inconsistency between pre-equated and post-equated results is more rather than less likely using a two and three parameter model.

Conclusion

Overall, the practice of reliance on pre-equated tests to construct score tables prior to an operational assessment was supported by this investigation. Even when item parameters were obtained from a field test involving off-grade students, given at a different time of year, with no stakes, the score tables produced were generally very consistent with those obtained by the full state-wide population in a high stakes operational assessment. Although measurement equivalence, calculated by DTF under the DFIT framework, was generally not established between the field test and the operational population, the true score to theta tables produced for each event were remarkably similar.

Importantly, a relationship was found between DTF and score table consistency. While no firm conclusion was established to indicate what degree of departure from measurement equivalence will lead to a problematic inconsistency in pre-equating, it remains a facet that requires close attention from test developers.

References

Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R.L. Thorndike (Ed.), *Educational Measurement* (2nd ed.) (pp. 508-600). Washington, DC: American Council on Education.

Ausubel, D. P., Sullivan, E.V., & Ives, S.W. (1980). *Theory and problems of child development*. Third Edition. New York: Grune & Stratton.

Baker, F. B. (1993). *Equate 2.1: Computer program for equating two metrics in item response theory* (Computer program). Madison: University of Wisconsin, Laboratory of Experimental Design.

Beard, J. G., Fletcher, G., & Richards, L. (1984, April). *Domain-referenced tests: Achieving Equivalence Through Rasch Pre-Equating.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Bolt, D. M. (2002). A Monte Carlo comparison of parametric and nonparametric polytomous DIF detection methods. *Applied Measurement in Education, 2*, 113-141.

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences.* Mahwah, NJ: Lawrence Erlbaum Associates.

Ceci, S. (1991). How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. *Developmental Psychology,* 27, 703-722.

Drasgow, F., & Kanfer, R. (1985). Equivalence of psychological measurement in heterogeneous populations. *Journal of Applied Psychology,* 70, 662-680.

Fleer, P. F. (1993). A monte carlo assessment of a new measure of item and test bias. (Doctoral dissertation, Illinois Institute of Technology). *Dissertation Abstracts International*, 54-04, 2266B.

Flowers, C. P., Oshima, T. C., & Raju, N. S. (1999). A description and demonstration of the polytomous-DFIT framework. *Applied Psychological Measurement*, *23*, 309-326.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.

Hanson, B. A. & Zeng, L. (1995). PIE: A computer program for IRT equating. (Computer program). Iowa City: American College Testing.

Kim, S.-H., & Cohen, A. S. (1991). A comparison of two area measures for detecting differential item functioning. *Applied Psychological Measurement, 15*, 269-278.

Kolen, M. J., & Brennan, R. L. (1995) *Test equating methods and practices.* New York: Springer-Verlag.

Linacre, J. M., & Wright, B. D. (2004). *Winsteps 3.49* (Computer program). Chicago, IL: MESA Press.

Livingston, S. A. (1985, April*). Large sample pre-equating: How accurate?* Paper

presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Matlin, M. W. (2002). *Cognition.* Fifth edition. Orlando, FL: Harcourt College Publishers.

Mislevy, R. J., & Bock, R. D. (1997). *Bilog 3.11: Item analysis and test scoring with binary logistic models.* (Computer program). Scientific Software International.

Oshima, T. C., Raju, N. S., & Nanda, A. O. (2006). A new method for assessing the statistical significance in the differential functioning of items and tests (DFIT) framework. *Journal of Educational Measurement,* 43, 1-17.

Raju, N. S. (1990). Determining the significance of estimated signed and unsigned areas between two item response functions. *Applied Psychological Measurement,* 14, 197-207.

Raju, N. S. (2005). *DFITD7: A fortran program for calculating DIF/DTF.* (Computer program). Illinois Institute of Technology.

Raju, N. S., Laffitte, L. J., & Byrne, B. M. (2002). Measurement equivalence: A comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology*, 87 (3), 517-529.

Raju, N. S., van der Linden, W., & Fleer, P. (1995). IRT-based internal measure of test bias with applications for differential item functioning. *Applied Psychological Measurement*, 19, 353-368.

Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests.

Copenhagen: Denmarks Paedagogiske Institut. (Republished in 1980 by the University of Chicago Press, Chicago, IL).

Rice, F. P. (1997). *Child and adolescent development.* Upper Saddle River, NJ: Prentice Hall.

Stocking, M. L., & Eignor, D. R. (1986). *The impact of different ability distributions on IRT preequating.* (Rep. No. ETS-RR-86-49). Princeton, NJ: Educational Testing Service.

Stocking, M. L. & Lord, F. M.. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement*, 7(2), 201-210.

Thissen, D., Steinberg, L., & Wainer, H. (1988). Use of item response theory in the study of group differences in trace lines. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp.147-169). Hillsdale, NJ: Erlbaum.

Weiner, B. (1990). History of motivational research in education. *Journal of Educational Psychology,* 82, 616-622.

Wolfe, E. W. (2001). Equating and item banking with the Rasch model. *Journal of Applied Measurement*, 1(4), 409-434.

Wolf, L. F., & Smith, J. K. (1995). The consequence of consequence: Motivation, anxiety, and test performance. *Applied Measurement in Education*, 8(3), 227-242.

APPENDIX A

PRE-EQUATED AND POST-EQUATED ITEM PARAMETERS

Grade 3, English/ Language Arts

Item	Pre-Equated Difficulty	Post-Equated Difficulty	Difference
1	-1.6496	-1.3873	-0.2623
2	0.003	-0.9019	0.9049
3*	-0.3009	-0.3009	0
4	0.5545	0.7203	-0.1658
5	-1.2067	-1.9198	0.7131
6*	-0.0392	-0.0392	0
7	0.2424	-0.1734	0.4158
8	-0.389	-0.6868	0.2978
9*	-0.2065	-0.2065	0
10	0.2819	0.5716	-0.2897
11	-0.0567	0.484	-0.5407
12*	0.2389	0.2389	0
13	-1.2666	-0.9863	-0.2803
14	-1.0585	-1.0053	-0.0532
15*	-0.4515	-0.4515	0
16	0.2527	0.3425	-0.0898
17	0.6022	0.9494	-0.3472
18*	0.1749	0.1749	0
19	-0.9741	-0.5184	-0.4557
20	0.0365	0.4134	-0.3769
21	-0.5463	-0.558	0.0117
22	0.8624	0.7843	0.0781
23	0.1792	0.3418	-0.1626
24	-0.4308	-0.4294	-0.0014
25	1.069	0.9869	0.0821
26	0.1885	0.6882	-0.4997
27	-0.8435	-1.0848	0.2413
28	-0.1026	-0.0892	-0.0134
29	0.5592	0.7356	-0.1764
30	0.2247	-0.201	0.4257
31	0.9074	1.6922	-0.7848
32*	0.2571	0.2571	0
33	1.8437	2.0463	-0.2026
34	-0.4345	-0.4297	-0.0048
35*	-0.8098	-0.8098	0

36	1.6007	1.8155	-0.2148
37*	-0.0176	-0.0176	0
38	1.018	1.0296	-0.0116
39	0.0274	-0.1529	0.1803
40*	0.03	0.03	0
41	0.2746	-0.0508	0.3254
42*	0.5621	0.5621	0
43	0.3005	0.3215	-0.021
44	-0.2267	-0.3228	0.0961
45*	0.5582	0.5582	0
46	0.0618	-0.069	0.1308
47	0.9644	1.4037	-0.4393
48	0.6616	0.6901	-0.0285
49	0.8954	0.9724	-0.077
50	0.1355	-0.3558	0.4913
		Root Mean Square Difference=**.2297**	

* = Link Item

Grade 3, Math

Item	Pre-Equated Difficulty	Post-Equated Difficulty	Difference
1*	-0.2287	-0.2287	0
2	-0.8461	-1.9824	1.1363
3	-0.6274	-0.694	0.0666
4	0.4802	0.2981	0.1821
5	0.9525	0.8001	0.1524
6	0.3438	0.0511	0.2927
7	-1.0166	-0.9704	-0.0462
8*	-0.3255	-0.3255	0
9	-2.0593	-2.2113	0.152
10	1.1967	0.8752	0.3215
11	-0.7799	-1.5755	0.7956
12*	0.3994	0.3994	0
13	0.3639	0.3482	0.0157
14	-1.1734	-1.9017	0.7283
15	-0.5613	-0.6124	0.0511
16*	0.9508	0.9508	0
17	-0.6543	-0.8314	0.1771
18	-0.4077	-1.5401	1.1324
19	0.673	0.6795	-0.0065
20*	1.5301	1.5301	0
21	-0.0611	0.9921	-1.0532
22	-0.5883	0.2543	-0.8426
23	-2.1732	-1.4094	-0.7638
24*	-1.0247	-1.0247	0
25	0.9347	1.1149	-0.1802
26	0.564	0.1172	0.4468
27	0.766	0.5831	0.1829
28	1.1033	0.9639	0.1394
29	0.1399	0.1802	-0.0403
30	1.5961	1.7233	-0.1272
31	0.9332	1.2462	-0.313
32	1.8819	2.1352	-0.2533
33	-0.5024	-0.3701	-0.1323
34*	-0.3521	-0.3521	0
35	-1.0593	-0.8423	-0.217
36	1.1868	1.1558	0.031
37	0.2458	-0.0655	0.3113
38	0.0117	-0.1824	0.1941
39	0.9322	1.1401	-0.2079
40	0.6833	0.8491	-0.1658
41*	-0.4689	-0.4689	0
42	-0.8901	-1.3312	0.4411
43	-0.9173	-1.4942	0.5769
44	-0.6574	-0.8972	0.2398
45*	0.2657	0.2657	0
46	-0.2131	0.2889	-0.502

47	0.3899	0.5215	-0.1316
48*	0.6558	0.6558	0
49	-1.5955	-0.8882	-0.7073
50	0.135	-0.3189	0.4539
51	-0.3657	-0.3456	-0.0201
52*	0.0046	0.0046	0
53	2.2036	2.3016	-0.098
54	-0.3816	-0.4399	0.0583
55*	0.4266	0.4266	0
56	-0.082	0.3458	-0.4278
57	0.0809	0.7643	-0.6834
58	1.4092	1.4916	-0.0824
59	-0.223	0.1243	-0.3473
60	-0.2279	-0.032	-0.1959
		Root Mean Square Difference=**.3995**	

Grade 3, Social Studies

Item	Pre-Equated Difficulty	Post-Equated Difficulty	Difference
1	0.8415	0.5435	0.298
2	0.6677	0.0328	0.6349
3	0.1904	0.2947	-0.1043
4*	0.4025	0.4025	0
5	-0.6968	-1.4123	0.7155
6	-0.3779	-0.8347	0.4568
7	-0.2086	-0.563	0.3544
8*	0.0089	0.0089	0
9	-0.2252	-0.8821	0.6569
10	0.3247	-0.2003	0.525
11*	-0.1414	-0.1414	0
12	0.1194	0.0748	0.0446
13	-0.186	-0.4417	0.2557
14	-0.4758	-0.9116	0.4358
15*	0.757	0.757	0
16	-0.3457	-0.3103	-0.0354
17	0.8642	0.7054	0.1588
18*	-0.0091	-0.0091	0
19	0.3431	0.1567	0.1864
20	-0.4672	-0.9497	0.4825
21	0.2807	0.2077	0.073
22*	-0.548	-0.548	0
23	0.9502	0.8475	0.1027
24	1.0338	1.0552	-0.0214
25*	0.0335	0.0335	0
26	-0.669	-1.6906	1.0216
27	-0.2251	-0.1863	-0.0388
28	-0.1192	0.3387	-0.4579
29	1.1883	1.0014	0.1869
30	0.8156	0.8869	-0.0713
31	0.332	0.18	0.152
32	0.5692	0.6245	-0.0553
33	-0.9119	-1.0357	0.1238
34*	-0.1503	-0.1503	0
35	0.9807	0.9698	0.0109
36	1.0064	1.0155	-0.0091
37	0.0357	-0.0939	0.1296
38*	-0.0625	-0.0625	0
39	-0.2599	-0.3402	0.0803
40	0.0436	-0.2462	0.2898
41*	0.3284	0.3284	0
42	1.068	0.7572	0.3108
43	0.793	0.9074	-0.1144
44	0.8204	0.6423	0.1781
45*	-0.3212	-0.3212	0
46	-1.648	-1.3259	-0.3221

47	0.0261	0.1016	-0.0755
48	-0.0655	-0.0655	0
49	-0.382	-0.4291	0.0471
50	-0.7329	-0.5861	-0.1468
51	0.2476	0.404	-0.1564
52*	0.1759	0.1759	0
53	-0.4755	-0.7404	0.2649
54	0.936	0.4776	0.4584
55*	-0.079	-0.079	0
56	-0.2185	-0.3863	0.1678
57	-1.5178	-1.3742	-0.1436
58	-0.2739	-0.2373	-0.0366
59	-0.5127	-0.9616	0.4489
60	-1.4741	-0.6802	-0.7939
		Root Mean Square Difference=**.3036**	

Grade 5, English/ Language Arts

Item	Pre-Equated Difficulty	Post-Equated Difficulty	Difference
1	0.5424	0.4693	0.0731
2	0.4896	0.73	-0.2404
3	-0.0386	-0.6712	0.6326
4	-1.0706	-1.1305	0.0599
5	-0.7903	-2.5275	1.7372
6	-1.3462	-1.4989	0.1527
7*	-0.2127	-0.2127	0
8	0.6947	0.1389	0.5558
9	0.9125	0.941	-0.0285
10*	-0.0149	-0.0149	0
11	0.3373	0.4078	-0.0705
12*	-1.2497	-1.2497	0
13	-1.3132	-1.1735	-0.1397
14	0.8611	0.9678	-0.1067
15*	1.0595	1.0595	0
16	-0.6546	-0.6464	-0.0082
17*	-0.882	-0.882	0
18	0.367	0.56	-0.193
19	-0.3627	-0.5293	0.1666
20	-0.1682	-0.1945	0.0263
21	1.2574	0.9794	0.278
22	0.3732	0.3851	-0.0119
23	0.725	0.9581	-0.2331
24	1.474	1.5686	-0.0946
25	0.5828	0.8401	-0.2573
26	1.1452	1.3643	-0.2191
27	-0.6192	-0.5232	-0.096
28	-0.4303	-0.5352	0.1049
29	0.5773	0.49	0.0873
30	0.0326	-0.0791	0.1117
31	0.5588	0.6384	-0.0796
32*	0.664	0.664	0
33	-0.7782	-0.6908	-0.0874
34	1.1301	1.1276	0.0025
35*	0.7456	0.7456	0
36	1.1299	1.2867	-0.1568
37*	-0.3244	-0.3244	0
38	0.0834	0.1192	-0.0358
39	0.4445	0.8807	-0.4362
40*	0.1202	0.1202	0
41	0.0815	0.2097	-0.1282
42*	0.3542	0.3542	0
43	0.0156	0.7351	-0.7195
44	0.0077	0.2176	-0.2099
45*	0.9384	0.9384	0
46	1.6253	1.4925	0.1328
47	-0.3354	-0.1411	-0.1943

48	-0.891	-0.3579	-0.5331
49	-0.7432	-0.5169	-0.2263
50	-0.5209	-0.3723	-0.1486
		Root Mean Square Difference=**.3297**	

Grade 5, Math

Item	Pre-Equated Difficulty	Post-Equated Difficulty	Difference
1	0.4082	-0.0212	0.4294
2	-0.1984	-0.3023	0.1039
3	-0.1198	-0.5271	0.4073
4*	-0.9749	-0.9749	0
5	1.0324	2.0168	-0.9844
6	-0.2049	-0.399	0.1941
7	-0.96	-1.0498	0.0898
8*	-0.5006	-0.5006	0
9	0.5406	0.1955	0.3451
10	-0.1117	-0.0205	-0.0912
11*	-0.4336	-0.4336	0
12	0.9175	0.6984	0.2191
13	0.6726	-0.0782	0.7508
14	-0.0702	0.1532	-0.2234
15*	-0.0491	-0.0491	0
16	1.2594	1.3679	-0.1085
17	0.7063	0.6693	0.037
18*	0.3815	0.3815	0
19	1.2098	1.1754	0.0344
20	-0.2302	-0.607	0.3768
21	-0.1277	-0.7172	0.5895
22*	0.5562	0.5562	0
23	0.1564	-0.0374	0.1938
24	-0.8716	-0.6943	-0.1773
25*	-0.508	-0.508	0
26	0.22	0.4197	-0.1997
27	0.3881	0.6506	-0.2625
28	-2.1649	-1.7767	-0.3882
29	0.1843	0.7418	-0.5575
30	0.2784	0.8829	-0.6045
31	-0.025	-0.3925	0.3675
32	-0.5777	-0.7227	0.145
33	-1.3168	-1.976	0.6592
34*	0.186	0.186	0
35	-0.7357	-0.8323	0.0966
36	1.1085	0.9379	0.1706
37	-0.2477	-0.4701	0.2224
38*	0.4894	0.4894	0
39	0.9402	1.0948	-0.1546
40	-0.6046	-0.3487	-0.2559
41*	-1.1403	-1.1403	0
42	-0.0394	0.137	-0.1764
43	-0.6797	-0.1962	-0.4835
44	-1.5966	-1.2047	-0.3919
45*	-0.4704	-0.4704	0
46	-0.9385	-1.3684	0.4299

47	-0.9208	-0.7339	-0.1869
48*	0.4108	0.4108	0
49	-0.0779	-0.0598	-0.0181
50	-0.7905	-0.2812	-0.5093
51	0.3246	0.7765	-0.4519
52*	0.2861	0.2861	0
53	-1.0204	-0.8634	-0.157
54	-0.8263	-1.1528	0.3265
55*	0.3872	0.3872	0
56	-0.527	-0.2591	-0.2679
57	0.8316	0.8275	0.0041
58	0.8208	0.9997	-0.1789
59	0.5668	0.5907	-0.0239
60	-1.0429	-0.5548	-0.4881
		Root Mean Square Difference=**.3167**	

Grade 5, Social Studies

Item	Pre-Equated Difficulty	Post-Equated Difficulty	Difference
1	1.6064	-1.4684	3.0748
2	-0.8655	-0.7996	-0.0659
3	-0.2666	-0.5641	0.2975
4*	0.6745	0.6745	0
5	-0.064	-0.7267	0.6627
6	-0.2808	-0.2387	-0.0421
7	-0.2728	-0.7233	0.4505
8*	-0.2149	-0.2149	0
9	-0.843	-1.3486	0.5056
10	-0.2091	-0.1361	-0.073
11*	0.4156	0.4156	0
12	1.5117	1.5556	-0.0439
13	-0.3109	0.3446	-0.6555
14	0.035	0.2317	-0.1967
15*	0.7886	0.7886	0
16	0.6661	0.5112	0.1549
17	-2.6982	-2.6093	-0.0889
18*	0.1123	0.1123	0
19	0.0355	-0.1185	0.154
20	0.7973	0.7216	0.0757
21	0.2319	0.1402	0.0917
22*	-0.8865	-0.8865	0
23	0.8454	0.7311	0.1143
24	1.4185	1.3977	0.0208
25*	0.4337	0.4337	0
26	0.958	1.2114	-0.2534
27	-0.0089	-0.1377	0.1288
28	1.1285	1.1651	-0.0366
29	0.3687	0.3939	-0.0252
30	0.1431	-0.2578	0.4009
31	1.3005	1.6858	-0.3853
32	0.5712	0.6557	-0.0845
33	1.1493	1.4059	-0.2566
34*	-0.0443	-0.0443	0
35	1.1531	1.4389	-0.2858
36	0.302	0.2723	0.0297
37	1.2132	1.4402	-0.227
38*	-0.2506	-0.2506	0
39	0.174	0.0031	0.1709
40	0.8403	1.2197	-0.3794
41*	0.4924	0.4924	0
42	0.9897	1.1665	-0.1768
43	0.1387	0.4558	-0.3171
44	0.7678	1.0261	-0.2583
45*	-0.797	-0.797	0

46	0.35	0.3784	-0.0284
47	-0.3311	-0.2895	-0.0416
48*	-0.2998	-0.2998	0
49	0.6525	0.864	-0.2115
50	-1.9286	-1.2254	-0.7032
51	0.3694	0.086	0.2834
52*	0.1534	0.1534	0
53	0.3105	0.3688	-0.0583
54	0.1587	0.5117	-0.353
55*	-1.0621	-1.0621	0
56	-0.0287	0.0401	-0.0688
57	-0.4019	-0.2083	-0.1936
58	-0.4699	-0.2929	-0.177
59	-0.3656	0.6598	-1.0254
60	-2.4102	-1.7858	-0.6244
		Root Mean Square Difference=**.4872**	

Grade 7, English/ Language Arts

Item	Pre-Equated Difficulty	Post-Equated Difficulty	Difference
1	-0.0898	-0.2541	0.1643
2	-1.3673	-2.1845	0.8172
3	-0.6377	-0.6632	0.0255
4	-0.3169	-0.4291	0.1122
5	-1.0898	-2.1335	1.0437
6	-0.4214	-1.811	1.3896
7*	-0.5566	-0.5566	0
8	0.6167	1.1409	-0.5242
9	0.6493	0.6139	0.0354
10*	0.7278	0.7278	0
11	-0.7956	-1.0763	0.2807
12*	1.0452	1.0452	0
13	-0.0704	0.1338	-0.2042
14	-0.1253	-0.2499	0.1246
15*	-0.7757	-0.7757	0
16	-0.7389	-0.5074	-0.2315
17*	0.8545	0.8545	0
18	-0.721	-0.72	-0.001
19	-0.7224	-0.9875	0.2651
20*	0.9029	0.9029	0
21	-1.0287	-0.9751	-0.0536
22	0.067	0.07	-0.003
23	-0.6882	-0.8982	0.21
24	1.1296	1.1081	0.0215
25	-0.1588	0.0396	-0.1984
26	0.0022	-0.0201	0.0223
27	0.2969	0.3127	-0.0158
28	-0.581	-0.5695	-0.0115
29	0.4021	0.6412	-0.2391
30	1.9001	1.6243	0.2758
31	-0.963	-0.7523	-0.2107
32*	0.6477	0.6477	0
33	-0.5194	-0.6936	0.1742
34	-0.1132	-0.1848	0.0716
35*	-0.3628	-0.3628	0
36	-0.9454	-0.7133	-0.2321
37*	0.1345	0.1345	0
38	0.3794	0.6662	-0.2868
39	1.2345	1.2438	-0.0093
40*	-0.2009	-0.2009	0
41	0.0076	0.2036	-0.196
42*	-0.46	-0.46	0
43	0.4288	0.4186	0.0102
44	0.0325	-0.0594	0.0919
45*	0.6319	0.6319	0
46	0.6717	0.7988	-0.1271

47	0.3884	0.2713	0.1171
48	0.5878	0.6902	-0.1024
49	0.9704	1.2268	-0.2564
50	1.5399	0.5758	0.9641
		Root Mean Square Difference=**.3401**	

Grade 7, Math

Item	Pre-Equated Difficulty	Post-Equated Difficulty	Difference
1	1.0858	0.636	0.4498
2	-0.1954	-0.4537	0.2583
3	-0.6217	-0.7145	0.0928
4*	0.2424	0.2424	0
5	-0.4465	-0.4068	-0.0397
6	-0.3446	-0.8405	0.4959
7	0.6864	0.6605	0.0259
8*	-0.4281	-0.4281	0
9	1.1332	1.4844	-0.3512
10	-0.5639	-0.1391	-0.4248
11*	-0.2575	-0.2575	0
12	-1.2089	-1.4017	0.1928
13	0.5378	0.73	-0.1922
14	-0.2404	-0.2127	-0.0277
15*	-0.3464	-0.3464	0
16	-0.4452	-0.9906	0.5454
17	-1.6442	-2.0717	0.4275
18*	0.0257	0.0257	0
19	1.0051	0.88	0.1251
20	-0.0833	-0.4631	0.3798
21	-1.077	-0.8309	-0.2461
22*	-0.826	-0.826	0
23	0.6076	0.8504	-0.2428
24	-0.1757	-0.1991	0.0234
25*	-0.1291	-0.1291	0
26	-0.7483	-0.7588	0.0105
27	0.5039	0.5284	-0.0245
28	-1.085	-0.7974	-0.2876
29	-0.3218	-0.302	-0.0198
30	-0.2576	-0.1882	-0.0694
31	0.3407	0.9687	-0.628
32	-1.4312	-1.2258	-0.2054
33	-0.0697	0.2537	-0.3234
34*	-1.017	-1.017	0
35	-0.9077	-0.7901	-0.1176
36	-0.4204	-0.3401	-0.0803
37	0.5415	0.8598	-0.3183
38*	0.1007	0.1007	0
39	0.6002	0.1934	0.4068
40	0.1482	0.1382	0.01
41*	-0.3377	-0.3377	0
42	-0.9222	-1.3223	0.4001
43	-0.0656	-0.2798	0.2142
44	2.2562	1.878	0.3782
45*	0.482	0.482	0
46	0.347	0.1613	0.1857

47	-0.0126	0.2009	-0.2135
48*	-0.0533	-0.0533	0
49	0.9083	0.9947	-0.0864
50	-1.0712	-0.5252	-0.546
51	0.9675	0.8204	0.1471
52*	0.7921	0.7921	0
53	0.7201	0.9998	-0.2797
54	-0.8966	-0.4097	-0.4869
55*	0.535	0.535	0
56	-0.6777	-0.4215	-0.2562
57	1.1907	-0.0036	1.1943
58	1.6062	2.1512	-0.545
59	1.4495	1.9609	-0.5114
60	1.3849	1.469	-0.0841
		Root Mean Square Difference=**.3085**	

Grade 7, Social Studies

Item	Pre-Equated Difficulty	Post-Equated Difficulty	Difference
1	0.2703	-0.1216	0.3919
2	0.5519	0.2429	0.309
3	0.0667	-0.3997	0.4664
4*	0.2097	0.2097	0
5	0.7777	0.5296	0.2481
6	0.2872	0.2421	0.0451
7	0.7723	0.9062	-0.1339
8*	-0.1096	-0.1096	0
9	-0.1413	0.2347	-0.376
10	-0.0035	-0.5497	0.5462
11*	0.6519	0.6519	0
12	0.4601	0.7945	-0.3344
13	0.0605	-0.1521	0.2126
14	-1.3902	-1.4269	0.0367
15*	-0.341	-0.341	0
16	0.1718	-0.0161	0.1879
17	0.5214	0.4183	0.1031
18*	0.0301	0.0301	0
19	-1.2378	-0.3962	-0.8416
20	0.0792	-0.1627	0.2419
21	0.1473	0.5214	-0.3741
22*	-1.5582	-1.5582	0
23	0.2167	0.4428	-0.2261
24	-0.6557	-0.2549	-0.4008
25*	-0.2566	-0.2566	0
26	-0.0707	-0.026	-0.0447
27	0.8904	0.4874	0.403
28	0.465	0.5343	-0.0693
29	-1.5478	-1.4064	-0.1414
30	-0.7777	-0.6497	-0.128
31	-0.4887	-0.5672	0.0785
32	-0.242	-0.299	0.057
33	-0.3079	0.2381	-0.546
34*	-0.6379	-0.6379	0
35	0.1397	0.5122	-0.3725
36	0.1571	0.1022	0.0549
37	0.1093	0.1324	-0.0231
38*	-0.8431	-0.8431	0
39	-0.1128	-0.098	-0.0148
40	-0.3577	-0.6439	0.2862
41*	0.0163	0.0163	0
42	0.6754	0.8977	-0.2223
43	0.4245	0.2211	0.2034
44	0.1617	0.1565	0.0052
45*	0.6363	0.6363	0
46	0.5665	0.723	-0.1565

47	0.133	0.6394	-0.5064
48*	0.7066	0.7066	0
49	0.4503	0.4968	-0.0465
50	-0.9223	-0.8794	-0.0429
51	-0.0479	-0.4543	0.4064
52*	0.3276	0.3276	0
53	-0.8041	-0.6864	-0.1177
54	-0.2781	0.1044	-0.3825
55	0.669	0.4582	0.2108
56	-0.0021	0.4683	-0.4704
57	-0.9003	-0.8351	-0.0652
58	-0.8247	-0.9197	0.095
59	0.2438	0.2964	-0.0526
60	0.8086	0.7027	0.1059
		Root Mean Square Difference=**.2596**	

APPENDIX B

PRE-EQUATED AND POST-EQUATED SCORING TABLES

Grade 3 ELA Score Table Comparison

	Pre-Equated		Post-Equated			
Raw Score	Theta	Scale Score	Theta	Scale Score	Theta Difference	SS Difference
0	-5.2638	168	-5.3095	166	0.0457	1
1	-4.0366	200	-4.0798	199	0.0432	1
2	-3.3103	219	-3.3502	218	0.0399	1
3	-2.8716	231	-2.9085	230	0.0369	1
4	-2.5506	239	-2.5847	239	0.0341	1
5	-2.2939	246	-2.3255	245	0.0316	1
6	-2.0779	252	-2.1069	251	0.029	1
7	-1.8898	257	-1.9165	256	0.0267	1
8	-1.722	261	-1.7464	261	0.0244	1
9	-1.5696	265	-1.5918	265	0.0222	1
10	-1.4294	269	-1.4493	269	0.0199	1
11	-1.2986	273	-1.3164	272	0.0178	0
12	-1.1758	276	-1.1914	275	0.0156	0
13	-1.0594	279	-1.0727	279	0.0133	0
14	-0.9483	282	-0.9595	281	0.0112	0
15	-0.8417	285	-0.8507	284	0.009	0
16	-0.739	287	-0.7456	287	0.0066	0
17	-0.6394	290	-0.6437	290	0.0043	0
18	-0.5424	293	-0.5444	292	0.002	0
19	-0.4476	295	-0.4473	295	-0.0003	0
20	-0.3547	297	-0.3519	298	-0.0028	0
21	-0.2633	300	-0.258	300	-0.0053	0
22	-0.173	302	-0.1651	303	-0.0079	0
23	-0.0836	305	-0.0731	305	-0.0105	0
24	0.0053	307	0.0185	307	-0.0132	0
25	0.0939	309	0.1097	310	-0.0158	0
26	0.1825	312	0.2011	312	-0.0186	0
27	0.2713	314	0.2927	315	-0.0214	-1
28	0.3606	316	0.3849	317	-0.0243	-1
29	0.4507	319	0.478	320	-0.0273	-1
30	0.5418	321	0.5722	322	-0.0304	-1
31	0.6345	324	0.6679	325	-0.0334	-1
32	0.7288	326	0.7655	327	-0.0367	-1
33	0.8253	329	0.8653	330	-0.04	-1
34	0.9244	331	0.9678	332	-0.0434	-1
35	1.0267	334	1.0735	335	-0.0468	-1

36	1.1327	337	1.183	338	-0.0503	-1
37	1.2431	340	1.297	341	-0.0539	-1
38	1.3588	343	1.4165	344	-0.0577	-2
39	1.4809	346	1.5425	348	-0.0616	-2
40	1.6108	349	1.6764	351	-0.0656	-2
41	1.7503	353	1.82	355	-0.0697	-2
42	1.9018	357	1.9758	359	-0.074	-2
43	2.0687	362	2.1471	364	-0.0784	-2
44	2.256	367	2.3388	369	-0.0828	-2
45	2.4712	372	2.5587	375	-0.0875	-2
46	2.7269	379	2.8193	381	-0.0924	-2
47	3.0471	387	3.1444	390	-0.0973	-3
48	3.4849	399	3.5875	402	-0.1026	-3
49	4.2105	418	4.3184	421	-0.1079	-3
50	5.4371	451	5.5489	454	-0.1118	-3

Root Mean Square Theta Difference = .0496

Grade 3 Math Score Table Comparison

	Pre-Equated		Post-Equated			
Raw Score	Theta	Scale Score	Theta	Scale Score	Theta Difference	SS Difference
0	-5.6494	177	-5.7561	175	0.1067	2
1	-4.4202	204	-4.5253	202	0.1051	2
2	-3.6913	220	-3.794	218	0.1027	2
3	-3.2507	230	-3.3505	228	0.0998	2
4	-2.9282	237	-3.0248	235	0.0966	2
5	-2.6706	243	-2.7637	241	0.0931	2
6	-2.4541	247	-2.5436	246	0.0895	2
7	-2.266	252	-2.3515	250	0.0855	2
8	-2.0985	255	-2.1801	254	0.0816	2
9	-1.9468	259	-2.0244	257	0.0776	2
10	-1.8076	262	-1.881	260	0.0734	2
11	-1.6783	265	-1.7476	263	0.0693	2
12	-1.5571	267	-1.6224	266	0.0653	1
13	-1.4428	270	-1.5039	268	0.0611	1
14	-1.3342	272	-1.3912	271	0.057	1
15	-1.2304	274	-1.2834	273	0.053	1
16	-1.1307	277	-1.1798	276	0.0491	1
17	-1.0346	279	-1.0798	278	0.0452	1
18	-0.9415	281	-0.9829	280	0.0414	1
19	-0.8511	283	-0.8887	282	0.0376	1
20	-0.763	285	-0.7969	284	0.0339	1
21	-0.6767	287	-0.7072	286	0.0305	1
22	-0.5922	288	-0.6192	288	0.027	1
23	-0.5092	290	-0.5328	290	0.0236	1
24	-0.4273	292	-0.4477	292	0.0204	0
25	-0.3465	294	-0.3636	294	0.0171	0
26	-0.2664	296	-0.2804	295	0.014	0
27	-0.187	297	-0.198	297	0.011	0
28	-0.108	299	-0.1162	299	0.0082	0
29	-0.0294	301	-0.0347	301	0.0053	0
30	0.0491	303	0.0466	303	0.0025	0
31	0.1276	304	0.1277	304	-0.0001	0
32	0.2064	306	0.209	306	-0.0026	0
33	0.2854	308	0.2906	308	-0.0052	0
34	0.3649	310	0.3725	310	-0.0076	0
35	0.4451	311	0.455	312	-0.0099	0
36	0.5261	313	0.5384	313	-0.0123	0
37	0.6082	315	0.6226	315	-0.0144	0
38	0.6915	317	0.708	317	-0.0165	0
39	0.7762	319	0.7949	319	-0.0187	0
40	0.8626	321	0.8833	321	-0.0207	0
41	0.951	322	0.9737	323	-0.0227	-1
42	1.0417	324	1.0663	325	-0.0246	-1
43	1.135	327	1.1616	327	-0.0266	-1
44	1.2314	329	1.2598	329	-0.0284	-1
45	1.3313	331	1.3616	332	-0.0303	-1

46	1.4353	333	1.4675	334	-0.0322	-1
47	1.5441	336	1.5781	336	-0.034	-1
48	1.6585	338	1.6943	339	-0.0358	-1
49	1.7797	341	1.8173	342	-0.0376	-1
50	1.909	344	1.9484	344	-0.0394	-1
51	2.0481	347	2.0894	348	-0.0413	-1
52	2.1995	350	2.2427	351	-0.0432	-1
53	2.3665	354	2.4117	355	-0.0452	-1
54	2.5541	358	2.6013	359	-0.0472	-1
55	2.7698	363	2.8191	364	-0.0493	-1
56	3.0263	368	3.0778	369	-0.0515	-1
57	3.3475	375	3.4013	376	-0.0538	-1
58	3.7865	385	3.8426	386	-0.0561	-1
59	4.5134	401	4.572	402	-0.0586	-1
60	5.7411	428	5.8016	429	-0.0605	-1

Root Mean Square Theta Difference = .0517

Grade 3 Social Studies Score Table Comparison

	Pre-Equated		Post-Equated			
Raw Score	Theta	Scale Score	Theta	Scale Score	Theta Difference	SS Difference
0	-5.4696	185	-5.5924	182	0.1228	3
1	-4.2466	213	-4.3701	210	0.1235	3
2	-3.5265	230	-3.6506	227	0.1241	3
3	-3.0941	240	-3.2186	237	0.1245	3
4	-2.7794	247	-2.9043	244	0.1249	3
5	-2.5293	253	-2.6543	250	0.125	3
6	-2.3199	258	-2.4449	255	0.125	3
7	-2.1386	262	-2.2635	259	0.1249	3
8	-1.9778	266	-2.1024	263	0.1246	3
9	-1.8325	269	-1.9569	266	0.1244	3
10	-1.6994	272	-1.8235	269	0.1241	3
11	-1.5762	275	-1.6999	272	0.1237	3
12	-1.4611	278	-1.5842	275	0.1231	3
13	-1.3526	280	-1.4751	277	0.1225	3
14	-1.2497	283	-1.3717	280	0.122	3
15	-1.1516	285	-1.273	282	0.1214	3
16	-1.0576	287	-1.1783	284	0.1207	3
17	-0.967	289	-1.0871	286	0.1201	3
18	-0.8795	291	-0.9988	288	0.1193	3
19	-0.7946	293	-0.9132	290	0.1186	3
20	-0.7119	295	-0.8298	292	0.1179	3
21	-0.6312	297	-0.7484	294	0.1172	3
22	-0.5522	299	-0.6686	296	0.1164	3
23	-0.4746	301	-0.5902	298	0.1156	3
24	-0.3982	302	-0.513	300	0.1148	3
25	-0.3228	304	-0.4369	301	0.1141	3
26	-0.2482	306	-0.3616	303	0.1134	3
27	-0.1743	308	-0.287	305	0.1127	3
28	-0.1009	309	-0.2128	307	0.1119	3
29	-0.0278	311	-0.1391	308	0.1113	3
30	0.0451	313	-0.0654	310	0.1105	3
31	0.118	314	0.0082	312	0.1098	3
32	0.1911	316	0.0819	313	0.1092	3
33	0.2644	318	0.1559	315	0.1085	3
34	0.3382	319	0.2303	317	0.1079	2
35	0.4125	321	0.3053	319	0.1072	2
36	0.4877	323	0.3811	320	0.1066	2
37	0.5638	325	0.4578	322	0.106	2
38	0.641	326	0.5356	324	0.1054	2
39	0.7197	328	0.6148	326	0.1049	2
40	0.7999	330	0.6956	328	0.1043	2
41	0.882	332	0.7782	330	0.1038	2
42	0.9663	334	0.8631	332	0.1032	2
43	1.0532	336	0.9504	334	0.1028	2
44	1.1429	338	1.0406	336	0.1023	2
45	1.2361	340	1.1343	338	0.1018	2

46	1.3332	342	1.2319	340	0.1013	2
47	1.435	345	1.3341	342	0.1009	2
48	1.5422	347	1.4418	345	0.1004	2
49	1.656	350	1.556	348	0.1	2
50	1.7777	353	1.6781	350	0.0996	2
51	1.9091	356	1.8098	353	0.0993	2
52	2.0525	359	1.9535	357	0.099	2
53	2.2113	363	2.1127	360	0.0986	2
54	2.3903	367	2.2921	365	0.0982	2
55	2.5971	372	2.4993	369	0.0978	2
56	2.8445	377	2.747	375	0.0975	2
57	3.1562	385	3.0589	382	0.0973	2
58	3.5852	395	3.4883	392	0.0969	2
59	4.3017	411	4.205	409	0.0967	2
60	5.5218	439	5.4254	437	0.0964	2
					Root Mean Square Theta Difference = .1156	

Grade 5 ELA Score Table Comparison

	Pre-Equated		Post-Equated			
Raw Score	Theta	Scale Score	Theta	Scale Score	Theta Difference	SS Difference
0	-5.2583	170	-5.3889	166	0.1306	3
1	-4.0314	200	-4.1463	197	0.1149	3
2	-3.3053	218	-3.4011	216	0.0958	2
3	-2.8666	229	-2.9466	227	0.08	2
4	-2.5452	237	-2.612	236	0.0668	2
5	-2.288	244	-2.3434	242	0.0554	1
6	-2.0712	249	-2.117	248	0.0458	1
7	-1.8821	254	-1.9195	253	0.0374	1
8	-1.7132	258	-1.7433	257	0.0301	1
9	-1.5595	262	-1.5831	261	0.0236	1
10	-1.4178	266	-1.4355	265	0.0177	0
11	-1.2855	269	-1.298	269	0.0125	0
12	-1.1609	272	-1.1688	272	0.0079	0
13	-1.0427	275	-1.0463	275	0.0036	0
14	-0.9297	278	-0.9295	278	-0.0002	0
15	-0.8212	280	-0.8174	281	-0.0038	0
16	-0.7164	283	-0.7093	283	-0.0071	0
17	-0.6147	286	-0.6046	286	-0.0101	0
18	-0.5156	288	-0.5027	288	-0.0129	0
19	-0.4186	291	-0.4031	291	-0.0155	0
20	-0.3235	293	-0.3056	293	-0.0179	0
21	-0.2298	295	-0.2096	296	-0.0202	-1
22	-0.1372	298	-0.1149	298	-0.0223	-1
23	-0.0455	300	-0.0212	300	-0.0243	-1
24	0.0456	302	0.0718	303	-0.0262	-1
25	0.1364	304	0.1644	305	-0.028	-1
26	0.2271	307	0.2568	307	-0.0297	-1
27	0.3182	309	0.3494	310	-0.0312	-1
28	0.4096	311	0.4423	312	-0.0327	-1
29	0.5018	314	0.5358	314	-0.034	-1
30	0.5951	316	0.6304	317	-0.0353	-1
31	0.6897	318	0.7262	319	-0.0365	-1
32	0.786	321	0.8236	322	-0.0376	-1
33	0.8844	323	0.9231	324	-0.0387	-1
34	0.9853	326	1.025	327	-0.0397	-1
35	1.0893	328	1.1298	329	-0.0405	-1
36	1.1968	331	1.2382	332	-0.0414	-1
37	1.3087	334	1.3509	335	-0.0422	-1
38	1.4258	337	1.4687	338	-0.0429	-1
39	1.5491	340	1.5926	341	-0.0435	-1
40	1.6801	343	1.7242	344	-0.0441	-1
41	1.8205	347	1.8651	348	-0.0446	-1
42	1.9727	350	2.0177	351	-0.045	-1
43	2.1401	355	2.1854	356	-0.0453	-1
44	2.3276	359	2.3733	360	-0.0457	-1
45	2.5428	365	2.5886	366	-0.0458	-1

46	2.7983	371	2.8443	372	-0.046	-1
47	3.1179	379	3.164	380	-0.0461	-1
48	3.5549	390	3.601	391	-0.0461	-1
49	4.2792	408	4.3252	409	-0.046	-1
50	5.5046	439	5.5506	440	-0.046	-1
					Root Mean Square Theta Difference = .0455	

Grade 5 Math Score Table Comparison

	Pre-Equated		Post-Equated			
Raw Score	Theta	Scale Score	Theta	Scale Score	Theta Difference	SS Difference
0	-5.672	165	-5.6714	165	-0.0006	0
1	-4.4469	196	-4.4469	196	0	0
2	-3.7239	214	-3.7245	214	0.0006	0
3	-3.2887	225	-3.2898	225	0.0011	0
4	-2.9715	233	-2.973	233	0.0015	0
5	-2.7188	239	-2.7205	239	0.0017	0
6	-2.5071	244	-2.5088	244	0.0017	0
7	-2.3234	249	-2.3251	249	0.0017	0
8	-2.1603	253	-2.1619	253	0.0016	0
9	-2.0128	257	-2.0142	257	0.0014	0
10	-1.8775	260	-1.8788	260	0.0013	0
11	-1.7521	263	-1.7531	263	0.001	0
12	-1.6348	266	-1.6354	266	0.0006	0
13	-1.5241	269	-1.5244	269	0.0003	0
14	-1.4191	272	-1.419	272	-1E-04	0
15	-1.3188	274	-1.3183	274	-0.0005	0
16	-1.2226	276	-1.2217	276	-0.0009	0
17	-1.13	279	-1.1285	279	-0.0015	0
18	-1.0403	281	-1.0383	281	-0.002	0
19	-0.9532	283	-0.9506	283	-0.0026	0
20	-0.8684	285	-0.8653	285	-0.0031	0
21	-0.7855	287	-0.7818	287	-0.0037	0
22	-0.7044	289	-0.7	290	-0.0044	0
23	-0.6246	291	-0.6196	292	-0.005	0
24	-0.5461	293	-0.5404	293	-0.0057	0
25	-0.4686	295	-0.4621	295	-0.0065	0
26	-0.392	297	-0.3847	297	-0.0073	0
27	-0.316	299	-0.3079	299	-0.0081	0
28	-0.2405	301	-0.2316	301	-0.0089	0
29	-0.1654	303	-0.1555	303	-0.0099	0
30	-0.0904	305	-0.0796	305	-0.0108	0
31	-0.0155	307	-0.0036	307	-0.0119	0
32	0.0596	308	0.0724	309	-0.0128	0
33	0.1349	310	0.1488	311	-0.0139	0
34	0.2106	312	0.2257	313	-0.0151	0
35	0.287	314	0.3032	315	-0.0162	0
36	0.3641	316	0.3815	317	-0.0174	0
37	0.4421	318	0.4609	319	-0.0188	0
38	0.5213	320	0.5414	321	-0.0201	-1
39	0.6019	322	0.6234	323	-0.0215	-1
40	0.684	324	0.707	325	-0.023	-1
41	0.768	326	0.7926	327	-0.0246	-1
42	0.8541	328	0.8804	329	-0.0263	-1
43	0.9428	331	0.9708	331	-0.028	-1
44	1.0344	333	1.0642	334	-0.0298	-1
45	1.1293	335	1.1611	336	-0.0318	-1

46	1.2283	338	1.2621	339	-0.0338	-1
47	1.3319	340	1.3677	341	-0.0358	-1
48	1.4409	343	1.479	344	-0.0381	-1
49	1.5565	346	1.597	347	-0.0405	-1
50	1.68	349	1.723	350	-0.043	-1
51	1.8131	352	1.8588	353	-0.0457	-1
52	1.9582	356	2.0069	357	-0.0487	-1
53	2.1187	360	2.1706	361	-0.0519	-1
54	2.2995	364	2.3547	366	-0.0552	-1
55	2.5082	370	2.567	371	-0.0588	-1
56	2.7573	376	2.8201	378	-0.0628	-2
57	3.0707	384	3.1378	385	-0.0671	-2
58	3.5015	395	3.5733	396	-0.0718	-2
59	4.2197	412	4.2967	414	-0.077	-2
60	5.4411	443	5.5221	445	-0.081	-2
					Root Mean Square Theta Difference = .0300	

Grade 5 Social Studies Score Table Comparison

	Pre-Equated		Post-Equated			
Raw Score	Theta	Scale Score	Theta	Scale Score	Theta Difference	SS Difference
0	-5.643	176	-5.5897	177	-0.0533	-1
1	-4.3974	206	-4.3523	208	-0.0451	-1
2	-3.6482	225	-3.6136	226	-0.0346	-1
3	-3.1904	236	-3.165	237	-0.0254	-1
4	-2.8536	244	-2.836	245	-0.0176	0
5	-2.5839	251	-2.573	251	-0.0109	0
6	-2.3574	256	-2.352	257	-0.0054	0
7	-2.161	261	-2.1599	261	-0.0011	0
8	-1.9867	266	-1.9893	265	0.0026	0
9	-1.8295	269	-1.8349	269	0.0054	0
10	-1.6857	273	-1.6933	273	0.0076	0
11	-1.5528	276	-1.562	276	0.0092	0
12	-1.4288	279	-1.4393	279	0.0105	0
13	-1.3123	282	-1.3236	282	0.0113	0
14	-1.2021	285	-1.2138	284	0.0117	0
15	-1.0972	287	-1.1092	287	0.012	0
16	-0.9969	290	-1.0088	289	0.0119	0
17	-0.9006	292	-0.9121	292	0.0115	0
18	-0.8077	294	-0.8188	294	0.0111	0
19	-0.7177	297	-0.7282	296	0.0105	0
20	-0.6302	299	-0.6401	299	0.0099	0
21	-0.545	301	-0.554	301	0.009	0
22	-0.4617	303	-0.4698	303	0.0081	0
23	-0.3801	305	-0.3871	305	0.007	0
24	-0.2998	307	-0.3058	307	0.006	0
25	-0.2208	309	-0.2256	309	0.0048	0
26	-0.1427	311	-0.1464	311	0.0037	0
27	-0.0654	313	-0.0679	313	0.0025	0
28	0.0112	314	0.0101	314	0.0011	0
29	0.0874	316	0.0876	316	-0.0002	0
30	0.1634	318	0.1649	318	-0.0015	0
31	0.2392	320	0.2421	320	-0.0029	0
32	0.3151	322	0.3193	322	-0.0042	0
33	0.3911	324	0.3968	324	-0.0057	0
34	0.4676	326	0.4747	326	-0.0071	0
35	0.5445	328	0.5531	328	-0.0086	0
36	0.6222	329	0.6323	330	-0.0101	0
37	0.7008	331	0.7124	332	-0.0116	0
38	0.7806	333	0.7936	334	-0.013	0
39	0.8616	335	0.8761	336	-0.0145	0
40	0.9442	337	0.9603	338	-0.0161	0
41	1.0287	339	1.0462	340	-0.0175	0
42	1.1153	342	1.1344	342	-0.0191	0
43	1.2044	344	1.225	344	-0.0206	-1
44	1.2964	346	1.3186	347	-0.0222	-1

45	1.3918	348	1.4156	349	-0.0238	-1
46	1.4911	351	1.5164	351	-0.0253	-1
47	1.5951	353	1.6221	354	-0.027	-1
48	1.7046	356	1.7331	357	-0.0285	-1
49	1.8206	359	1.8506	360	-0.03	-1
50	1.9445	362	1.9762	363	-0.0317	-1
51	2.0781	365	2.1113	366	-0.0332	-1
52	2.2237	369	2.2585	370	-0.0348	-1
53	2.3846	373	2.4211	374	-0.0365	-1
54	2.5659	377	2.6039	378	-0.038	-1
55	2.775	382	2.8146	383	-0.0396	-1
56	3.0247	388	3.0659	389	-0.0412	-1
57	3.3386	396	3.3813	397	-0.0427	-1
58	3.77	407	3.8143	408	-0.0443	-1
59	4.4888	424	4.5346	425	-0.0458	-1
60	5.7106	454	5.7574	455	-0.0468	-1
			Root Mean Square Theta Difference = .0231			

Grade 7 ELA Score Table Comparison

	Pre-Equated		Post-Equated			
Raw Score	Theta	Scale Score	Theta	Scale Score	Theta Difference	SS Difference
0	-5.3178	164	-5.5361	159	0.2183	6
1	-4.0932	195	-4.3003	190	0.2071	5
2	-3.3703	214	-3.5624	209	0.1921	5
3	-2.9347	225	-3.113	220	0.1783	5
4	-2.6163	233	-2.7821	229	0.1658	4
5	-2.362	239	-2.5163	236	0.1543	4
6	-2.148	245	-2.2918	241	0.1438	4
7	-1.9616	250	-2.0957	246	0.1341	3
8	-1.7953	254	-1.9205	251	0.1252	3
9	-1.6441	258	-1.7612	255	0.1171	3
10	-1.5047	261	-1.6143	259	0.1096	3
11	-1.3747	265	-1.4773	262	0.1026	3
12	-1.2523	268	-1.3485	265	0.0962	2
13	-1.1361	271	-1.2263	268	0.0902	2
14	-1.0251	274	-1.1098	271	0.0847	2
15	-0.9184	276	-0.9978	274	0.0794	2
16	-0.8153	279	-0.8898	277	0.0745	2
17	-0.7151	282	-0.7852	280	0.0701	2
18	-0.6175	284	-0.6833	282	0.0658	2
19	-0.5219	286	-0.5837	285	0.0618	2
20	-0.428	289	-0.4861	287	0.0581	1
21	-0.3354	291	-0.39	290	0.0546	1
22	-0.2439	294	-0.2951	292	0.0512	1
23	-0.153	296	-0.2012	295	0.0482	1
24	-0.0626	298	-0.1079	297	0.0453	1
25	0.0275	300	-0.0151	299	0.0426	1
26	0.1178	303	0.0778	302	0.04	1
27	0.2083	305	0.1707	304	0.0376	1
28	0.2995	307	0.2641	306	0.0354	1
29	0.3916	310	0.3583	309	0.0333	1
30	0.4848	312	0.4534	311	0.0314	1
31	0.5795	315	0.5499	314	0.0296	1
32	0.6761	317	0.648	316	0.0281	1
33	0.7748	320	0.7482	319	0.0266	1
34	0.8762	322	0.8509	321	0.0253	1
35	0.9807	325	0.9566	324	0.0241	1
36	1.0891	328	1.0659	327	0.0232	1
37	1.2018	330	1.1795	330	0.0223	1
38	1.32	333	1.2983	333	0.0217	1
39	1.4446	337	1.4233	336	0.0213	1
40	1.5769	340	1.556	339	0.0209	1
41	1.7189	344	1.698	343	0.0209	1
42	1.8729	348	1.8519	347	0.021	1
43	2.0423	352	2.0209	351	0.0214	1
44	2.2321	357	2.21	356	0.0221	1

45	2.4497	362	2.4268	362	0.0229	1
46	2.708	369	2.6839	368	0.0241	1
47	3.0306	377	3.0051	376	0.0255	1
48	3.471	388	3.4437	388	0.0273	1
49	4.1989	407	4.1695	406	0.0294	1
50	5.4273	438	5.3962	437	0.0311	1
					Root Mean Square Theta Difference = .0868	

Grade 7 Math Score Table Comparison

	Pre-Equated		Post-Equated			
Raw Score	Theta	Scale Score	Theta	Scale Score	Theta Difference	SS Difference
0	-5.5726	160	-5.5955	160	0.0229	1
1	-4.3499	193	-4.3705	193	0.0206	1
2	-3.6299	213	-3.6475	212	0.0176	0
3	-3.1974	224	-3.2123	224	0.0149	0
4	-2.8824	233	-2.895	233	0.0126	0
5	-2.6316	240	-2.6422	239	0.0106	0
6	-2.4214	245	-2.4302	245	0.0088	0
7	-2.2391	250	-2.2463	250	0.0072	0
8	-2.0771	255	-2.0828	255	0.0057	0
9	-1.9304	259	-1.935	259	0.0046	0
10	-1.7959	262	-1.7993	262	0.0034	0
11	-1.671	266	-1.6735	266	0.0025	0
12	-1.5539	269	-1.5555	269	0.0016	0
13	-1.4435	272	-1.4443	272	0.0008	0
14	-1.3385	275	-1.3386	275	1E-04	0
15	-1.2382	277	-1.2375	277	-0.0007	0
16	-1.1418	280	-1.1406	280	-0.0012	0
17	-1.0488	282	-1.047	283	-0.0018	0
18	-0.9588	285	-0.9564	285	-0.0024	0
19	-0.8712	287	-0.8683	287	-0.0029	0
20	-0.7858	290	-0.7825	290	-0.0033	0
21	-0.7023	292	-0.6984	292	-0.0039	0
22	-0.6203	294	-0.6161	294	-0.0042	0
23	-0.5397	296	-0.535	296	-0.0047	0
24	-0.4602	298	-0.4551	299	-0.0051	0
25	-0.3817	300	-0.3761	301	-0.0056	0
26	-0.3038	303	-0.2978	303	-0.006	0
27	-0.2266	305	-0.2201	305	-0.0065	0
28	-0.1498	307	-0.1428	307	-0.007	0
29	-0.0731	309	-0.0657	309	-0.0074	0
30	0.0034	311	0.0113	311	-0.0079	0
31	0.0799	313	0.0884	313	-0.0085	0
32	0.1567	315	0.1657	315	-0.009	0
33	0.2339	317	0.2435	317	-0.0096	0
34	0.3115	319	0.3217	320	-0.0102	0
35	0.3899	321	0.4008	322	-0.0109	0
36	0.4692	323	0.4807	324	-0.0115	0
37	0.5495	326	0.5618	326	-0.0123	0
38	0.6311	328	0.6442	328	-0.0131	0
39	0.7142	330	0.7281	330	-0.0139	0
40	0.7989	332	0.8138	333	-0.0149	0
41	0.8857	335	0.9015	335	-0.0158	0
42	0.9748	337	0.9917	338	-0.0169	0
43	1.0666	340	1.0846	340	-0.018	0
44	1.1614	342	1.1805	343	-0.0191	-1

45	1.2598	345	1.2802	345	-0.0204	-1
46	1.3623	348	1.3841	348	-0.0218	-1
47	1.4696	351	1.4929	351	-0.0233	-1
48	1.5827	354	1.6074	354	-0.0247	-1
49	1.7025	357	1.7288	358	-0.0263	-1
50	1.8304	360	1.8585	361	-0.0281	-1
51	1.9682	364	1.9982	365	-0.03	-1
52	2.1184	368	2.1503	369	-0.0319	-1
53	2.2843	373	2.3181	373	-0.0338	-1
54	2.4708	378	2.5068	379	-0.036	-1
55	2.6856	383	2.7237	384	-0.0381	-1
56	2.9413	390	2.9816	391	-0.0403	-1
57	3.2618	399	3.3043	400	-0.0425	-1
58	3.7004	411	3.745	412	-0.0446	-1
59	4.427	430	4.4738	432	-0.0468	-1
60	5.6548	464	5.7029	465	-0.0481	-1

Root Mean Square Theta Difference = .0199

Grade 7 Social Studies Score Table Comparison

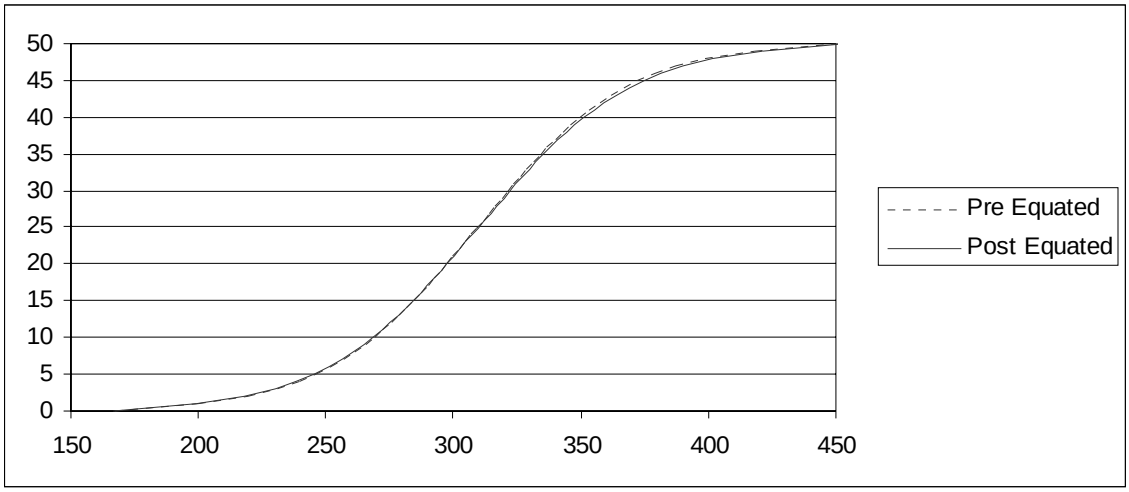
	Pre-Equated		Post-Equated			
Raw Score	Theta	Scale Score	Theta	Scale Score	Theta Difference	SS Difference
0	-5.5271	145	-5.4941	146	-0.033	-1
1	-4.3047	183	-4.2721	184	-0.0326	-1
2	-3.585	206	-3.5533	207	-0.0317	-1
3	-3.1531	219	-3.122	220	-0.0311	-1
4	-2.8388	229	-2.8083	230	-0.0305	-1
5	-2.5889	237	-2.5591	238	-0.0298	-1
6	-2.3797	244	-2.3506	245	-0.0291	-1
7	-2.1985	249	-2.17	250	-0.0285	-1
8	-2.0378	254	-2.0099	255	-0.0279	-1
9	-1.8927	259	-1.8652	260	-0.0275	-1
10	-1.7597	263	-1.7328	264	-0.0269	-1
11	-1.6366	267	-1.6102	268	-0.0264	-1
12	-1.5215	271	-1.4956	272	-0.0259	-1
13	-1.4131	274	-1.3876	275	-0.0255	-1
14	-1.3103	277	-1.2853	278	-0.025	-1
15	-1.2123	280	-1.1877	281	-0.0246	-1
16	-1.1184	283	-1.0942	284	-0.0242	-1
17	-1.0279	286	-1.0041	287	-0.0238	-1
18	-0.9406	289	-0.9171	290	-0.0235	-1
19	-0.8559	292	-0.8327	292	-0.0232	-1
20	-0.7735	294	-0.7506	295	-0.0229	-1
21	-0.6931	297	-0.6705	297	-0.0226	-1
22	-0.6144	299	-0.592	300	-0.0224	-1
23	-0.5372	302	-0.515	302	-0.0222	-1
24	-0.4612	304	-0.4392	305	-0.022	-1
25	-0.3863	306	-0.3645	307	-0.0218	-1
26	-0.3123	309	-0.2906	309	-0.0217	-1
27	-0.2389	311	-0.2174	312	-0.0215	-1
28	-0.1661	313	-0.1447	314	-0.0214	-1
29	-0.0937	316	-0.0724	316	-0.0213	-1
30	-0.0215	318	-0.0004	319	-0.0211	-1
31	0.0506	320	0.0717	321	-0.0211	-1
32	0.1228	322	0.1438	323	-0.021	-1
33	0.1952	325	0.2161	325	-0.0209	-1
34	0.2681	327	0.2889	328	-0.0208	-1
35	0.3414	329	0.3623	330	-0.0209	-1
36	0.4155	332	0.4363	332	-0.0208	-1
37	0.4905	334	0.5113	335	-0.0208	-1
38	0.5666	336	0.5874	337	-0.0208	-1
39	0.6441	339	0.6648	339	-0.0207	-1
40	0.723	341	0.7438	342	-0.0208	-1
41	0.8039	344	0.8246	344	-0.0207	-1
42	0.8868	346	0.9076	347	-0.0208	-1
43	0.9722	349	0.993	350	-0.0208	-1
44	1.0605	352	1.0814	353	-0.0209	-1

45	1.1522	355	1.173	355	-0.0208	-1
46	1.2478	358	1.2686	358	-0.0208	-1
47	1.348	361	1.3688	362	-0.0208	-1
48	1.4535	364	1.4745	365	-0.021	-1
49	1.5657	368	1.5866	368	-0.0209	-1
50	1.6856	372	1.7067	372	-0.0211	-1
51	1.8152	376	1.8363	376	-0.0211	-1
52	1.9567	380	1.9779	381	-0.0212	-1
53	2.1137	385	2.1349	386	-0.0212	-1
54	2.2909	391	2.3121	391	-0.0212	-1
55	2.4958	397	2.5171	398	-0.0213	-1
56	2.7412	405	2.7625	405	-0.0213	-1
57	3.0508	414	3.0722	415	-0.0214	-1
58	3.4779	428	3.4993	429	-0.0214	-1
59	4.1923	450	4.2137	451	-0.0214	-1
60	5.4109	489	5.4324	489	-0.0215	-1
					Root Mean Square Theta Difference = .0236	

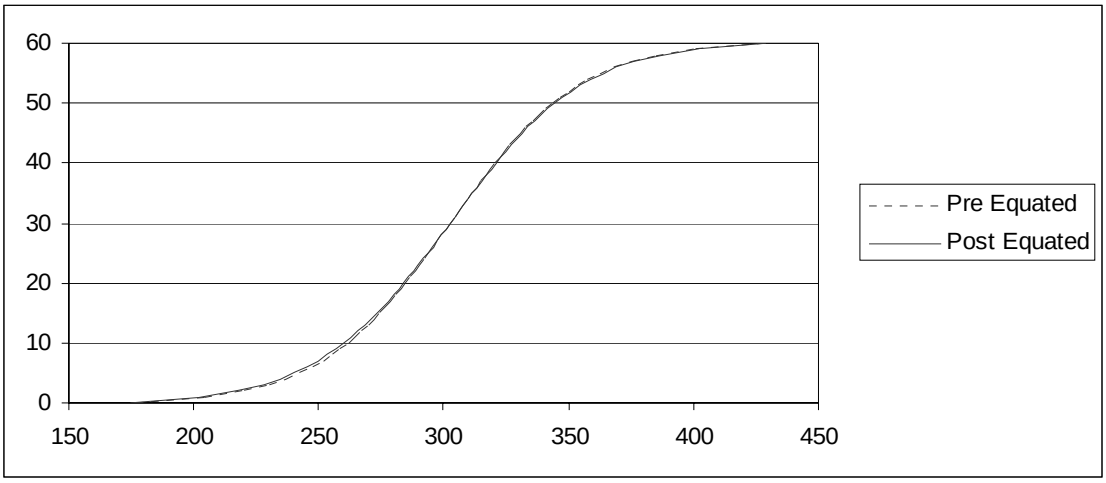
APPENDIX C

PRE-EQUATED AND POST-EQUATED SCORING GRAPHS

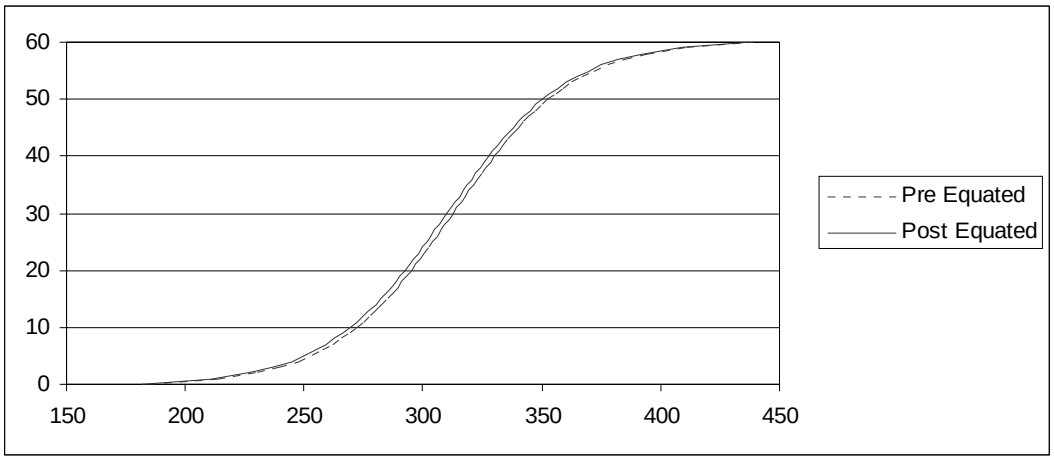
Grade 3 ELA



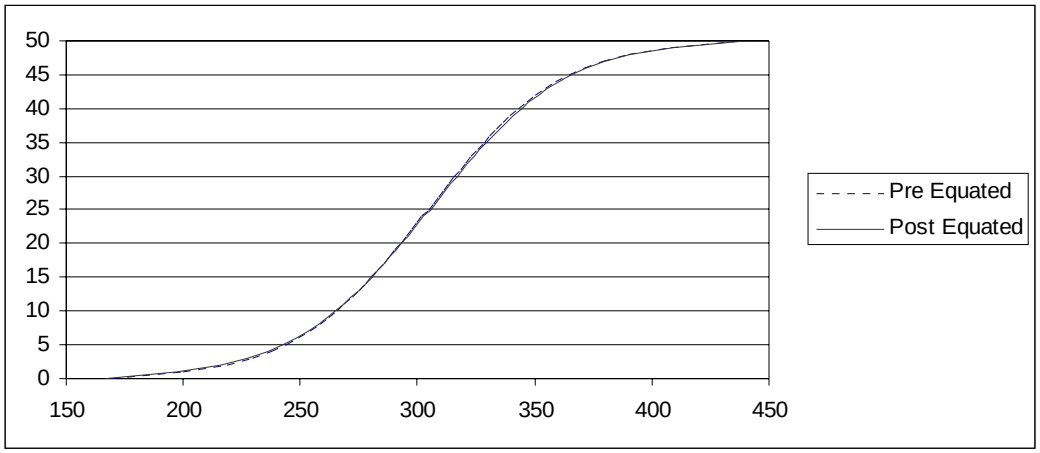
Grade 3 Math



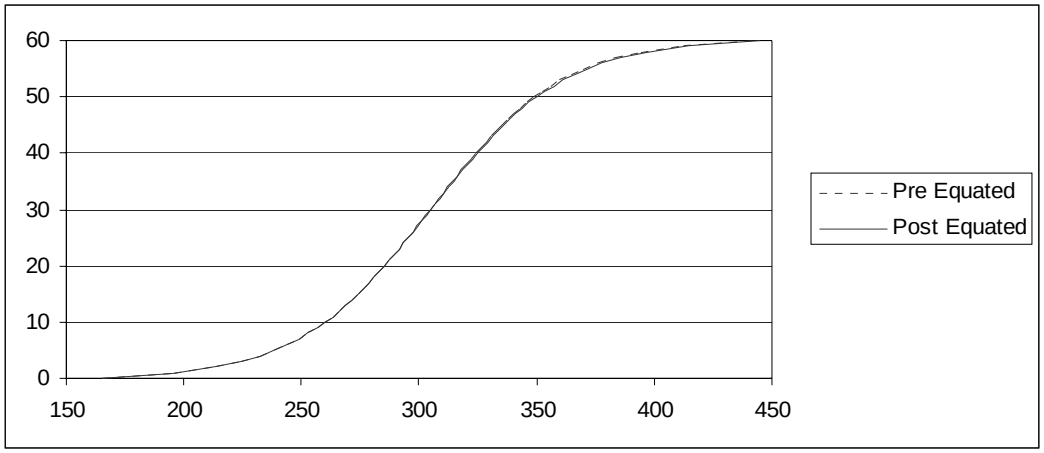
Grade 3 Social Studies



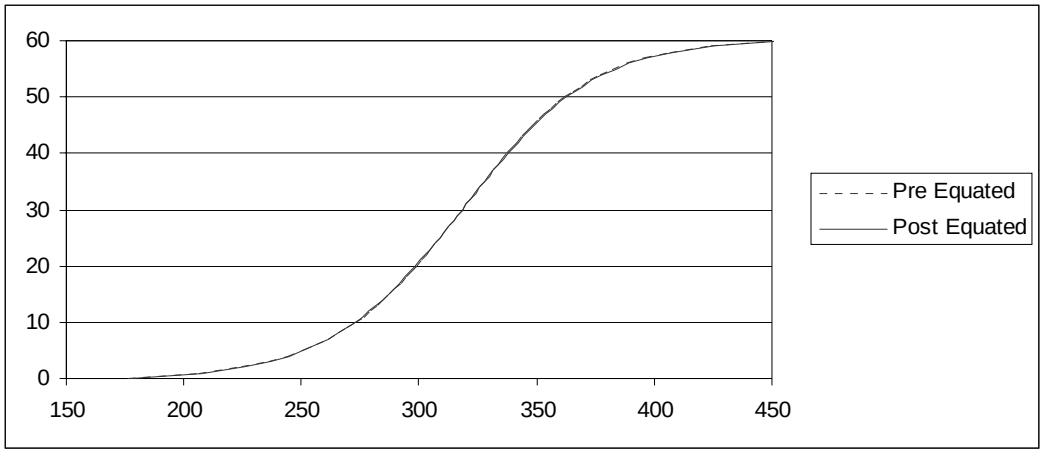
Grade 5 ELA



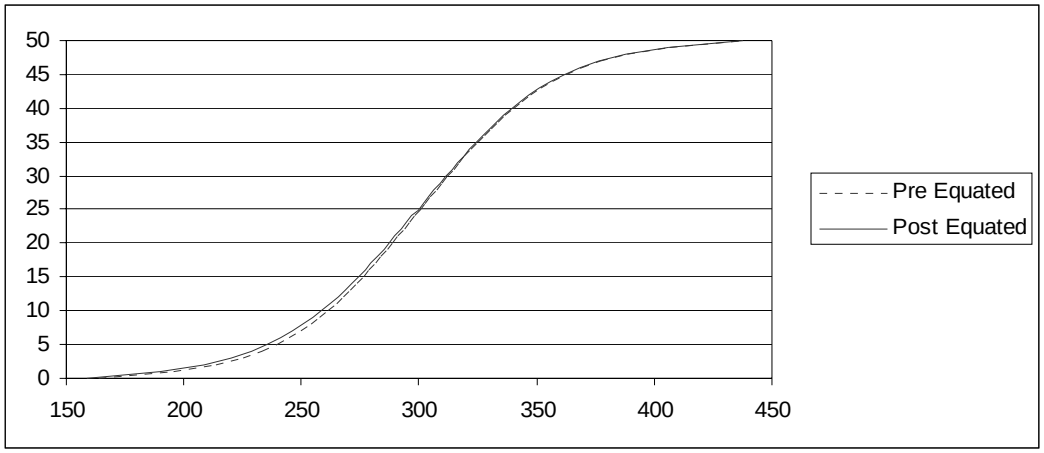
Grade 5 Math



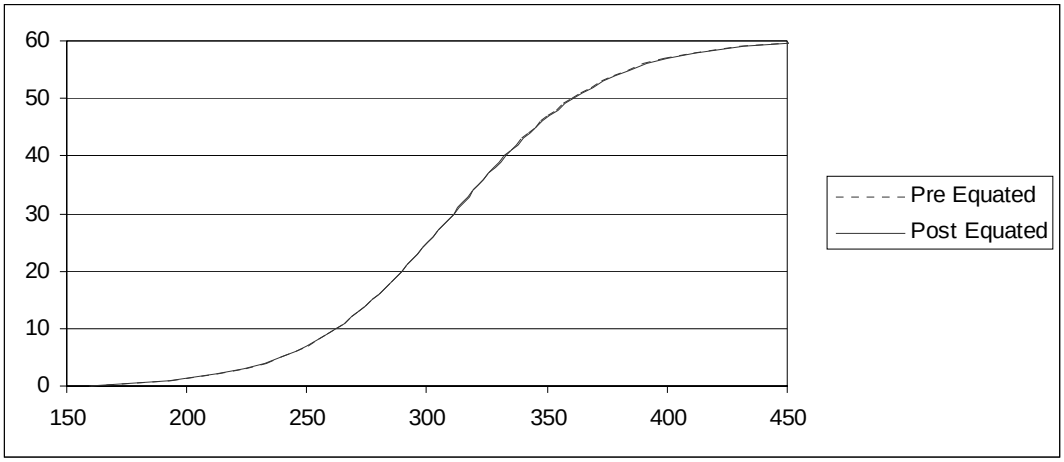
Grade 5 Social Studies



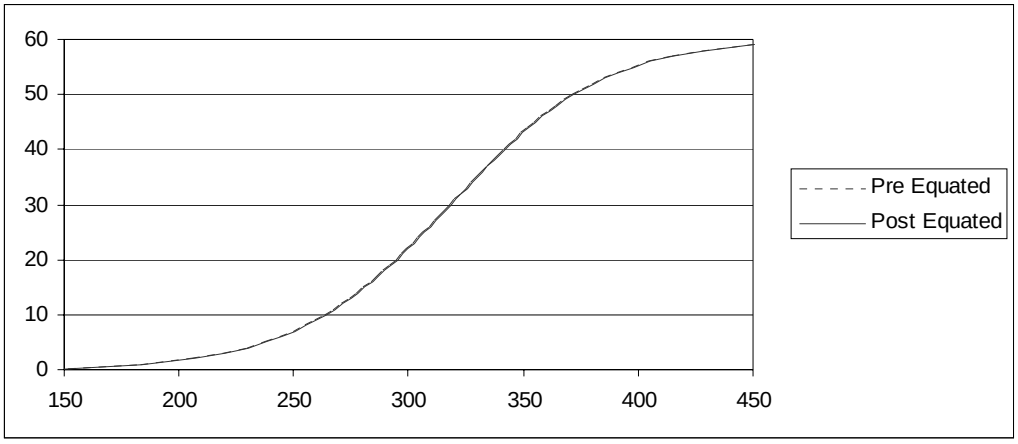
Grade 7 ELA



Grade 7 Math



Grade 7 Social Studies



APPENDIX D

PRELIMINARY DFIT ANALYSES WITH RANDOM GROUPS

Results for Random Group n=5000

Item	Mean(d)	SD(d)	CDIF	NCDIF	Sig.
1	0.0060	0.0090	0.0000	0.0001	ns
2	-0.0060	0.0070	0.0000	0.0001	ns
3	-0.0050	0.0040	0.0000	0.0000	ns
4	0.0200	0.0100	0.0001	0.0005	0.05
5	0.0010	0.0020	0.0000	0.0000	ns
6	0.0100	0.0070	0.0001	0.0001	ns
7	0.0080	0.0060	0.0001	0.0001	ns
8	0.0050	0.0050	0.0000	0.0001	ns
9	-0.0170	0.0130	-0.0001	0.0005	0.05
10	0.0080	0.0040	0.0001	0.0001	ns
11	0.0000	0.0000	0.0000	0.0000	ns
12	-0.0060	0.0040	-0.0001	0.0001	ns
13	-0.0090	0.0110	-0.0001	0.0002	ns
14	0.0000	0.0000	0.0000	0.0000	ns
15	0.0010	0.0000	0.0000	0.0000	ns
16	0.0070	0.0040	0.0001	0.0001	ns
17	0.0030	0.0020	0.0000	0.0000	ns
18	0.0100	0.0070	0.0001	0.0001	ns
19	0.0020	0.0020	0.0000	0.0000	ns
20	-0.0050	0.0030	0.0000	0.0000	ns
21	0.0030	0.0030	0.0000	0.0000	ns
22	0.0010	0.0010	0.0000	0.0000	ns
23	0.0020	0.0010	0.0000	0.0000	ns
24	0.0010	0.0010	0.0000	0.0000	ns
25	0.0040	0.0020	0.0000	0.0000	ns
26	0.0180	0.0090	0.0001	0.0004	0.05
27	-0.0050	0.0060	0.0000	0.0001	ns
28	-0.0070	0.0050	-0.0001	0.0001	ns
29	0.0100	0.0050	0.0001	0.0001	ns
30	-0.0090	0.0070	-0.0001	0.0001	ns
31	-0.0050	0.0030	0.0000	0.0000	ns
32	-0.0050	0.0030	0.0000	0.0000	ns
33	-0.0030	0.0020	0.0000	0.0000	ns
34	-0.0100	0.0090	-0.0001	0.0002	ns
35	0.0040	0.0040	0.0000	0.0000	ns
36	-0.0020	0.0010	0.0000	0.0000	ns

37	0.0050	0.0040	0.0000	0.0000	ns
38	-0.0080	0.0040	0.0000	0.0001	ns
39	0.0060	0.0040	0.0001	0.0001	ns
40	-0.0050	0.0040	0.0000	0.0000	ns
41	0.0020	0.0010	0.0000	0.0000	ns
42	-0.0100	0.0050	-0.0001	0.0001	ns
43	-0.0050	0.0030	0.0000	0.0000	ns
44	0.0060	0.0050	0.0001	0.0001	ns
45	-0.0170	0.0080	-0.0001	0.0004	ns
46	0.0080	0.0050	0.0001	0.0001	ns
47	-0.0240	0.0120	0.0001	0.0007	0.01
48	0.0060	0.0030	0.0000	0.0001	ns
49	0.0080	0.0040	0.0000	0.0001	ns
50	-0.0040	0.0030	0.0000	0.0000	ns
			Differential test functioning (DTF):		0.00028

Results for Random Group n=1000

Item	Mean(d)	SD(d)	CDIF	NCDIF	Sig.
1	0.0040	0.0060	0.0006	0.0001	ns
2	0.0260	0.0280	0.0030	0.0015	0.05
3	-0.0200	0.0160	-0.0014	0.0006	ns
4	0.0330	0.0170	-0.0008	0.0014	ns
5	-0.0040	0.0060	-0.0005	0.0001	ns
6	0.0260	0.0170	0.0010	0.0010	ns
7	-0.0010	0.0010	-0.0001	0.0000	ns
8	0.0070	0.0060	0.0006	0.0001	ns
9	0.0330	0.0250	0.0021	0.0017	0.05
10	0.0020	0.0010	-0.0001	0.0000	ns
11	-0.0010	0.0000	0.0000	0.0000	ns
12	0.0000	0.0000	0.0000	0.0000	ns
13	-0.0150	0.0180	-0.0018	0.0005	ns
14	0.0190	0.0230	0.0024	0.0009	ns
15	-0.0150	0.0120	-0.0010	0.0004	ns
16	-0.0250	0.0140	0.0000	0.0008	ns
17	-0.0160	0.0080	0.0010	0.0003	ns
18	0.0200	0.0150	0.0012	0.0006	ns
19	0.0170	0.0140	0.0014	0.0005	ns
20	-0.0070	0.0040	0.0000	0.0001	ns
21	-0.0080	0.0070	-0.0008	0.0001	ns
22	0.0380	0.0190	-0.0020	0.0018	0.05
23	-0.0150	0.0090	-0.0001	0.0003	ns
24	0.0020	0.0020	0.0002	0.0000	ns
25	-0.0060	0.0030	0.0004	0.0000	ns
26	0.0280	0.0150	-0.0008	0.0010	ns
27	-0.0060	0.0070	-0.0007	0.0001	ns
28	-0.0240	0.0170	-0.0011	0.0009	ns
29	0.0070	0.0030	-0.0003	0.0001	ns
30	0.0030	0.0020	0.0001	0.0000	ns
31	-0.0010	0.0010	0.0002	0.0000	ns
32	0.0210	0.0120	0.0003	0.0006	ns
33	0.0040	0.0030	-0.0005	0.0000	ns
34	-0.0190	0.0160	-0.0016	0.0006	ns
35	0.0060	0.0060	0.0006	0.0001	ns
36	-0.0650	0.0390	0.0071	0.0057	0.001
37	0.0020	0.0020	0.0001	0.0000	ns
38	-0.0700	0.0340	0.0039	0.0061	0.001
39	0.0520	0.0400	0.0033	0.0043	0.001
40	-0.0320	0.0210	-0.0013	0.0015	ns
41	-0.0100	0.0070	-0.0005	0.0001	ns
42	-0.0020	0.0010	0.0001	0.0000	ns
43	0.0000	0.0000	0.0000	0.0000	ns
44	0.0180	0.0140	0.0013	0.0005	ns
45	-0.0300	0.0150	0.0012	0.0011	ns
46	0.0360	0.0250	0.0018	0.0019	0.05

47	-0.0360	0.0190	0.0032	0.0016	0.05
48	0.0100	0.0050	-0.0003	0.0001	ns
49	-0.0010	0.0000	0.0001	0.0000	ns
50	-0.0160	0.0120	-0.0011	0.0004	ns
				Differential test functioning (DTF):	.0205

APPENDIX E

COMPARISON OF ITEM DIFFICULTY ESTIMATES FOR RANDOM SAMPLE OF 1000 TO FULL POPULATION OF 2004 EXAMINEES

Grade 3 ELA	
1000 Random	**Population**
-1.5524	-1.3873
-0.9870	-0.9019
-0.3009	-0.3009
0.5593	0.7203
-1.8813	-1.9198
-0.0392	-0.0392
-0.1469	-0.1734
-0.5745	-0.6868
-0.2065	-0.2065
0.5987	0.5716
0.4228	0.4840
0.2389	0.2389
-0.9611	-0.9863
-1.1823	-1.0053
-0.4515	-0.4515
0.4464	0.3425
1.0681	0.9494
0.1749	0.1749
-0.5271	-0.5184
0.4152	0.4134
-0.6065	-0.5580
0.8309	0.7843
0.3402	0.3418
-0.6086	-0.4294
1.0150	0.9869
0.5939	0.6882
-1.0125	-1.0848
0.0119	-0.0892
0.7919	0.7356
-0.0503	-0.2010
1.6331	1.6922
0.2571	0.2571
1.9998	2.0463
-0.3539	-0.4297
-0.8098	-0.8098
1.8926	1.8155

-0.0176	-0.0176
1.0675	1.0296
-0.3763	-0.1529
0.0300	0.0300
-0.0401	-0.0508
0.5621	0.5621
0.3201	0.3215
-0.4248	-0.3228
0.5582	0.5582
-0.1933	-0.0690
1.4395	1.4037
0.6470	0.6901
0.9599	0.9724
-0.2730	-0.3558
	Correlation=.99519

Grade 3 Math	
1000 Random	Population
-0.2287	-0.2287
-1.9031	-1.9824
-0.7636	-0.6940
0.3111	0.2981
0.6973	0.8001
0.1802	0.0511
-1.0442	-0.9704
-0.3255	-0.3255
-2.2732	-2.2113
0.8159	0.8752
-1.6208	-1.5755
0.3994	0.3994
0.3434	0.3482
-2.0638	-1.9017
-0.5693	-0.6124
0.9508	0.9508
-0.8072	-0.8314
-1.5918	-1.5401
0.7134	0.6795
1.5301	1.5301
0.9317	0.9921
0.0577	0.2543
-1.3182	-1.4094
-1.0247	-1.0247
1.2491	1.1149
0.1687	0.1172
0.6205	0.5831
0.9068	0.9639
0.3883	0.1802
1.8612	1.7233
1.3972	1.2462
2.1479	2.1352
-0.3851	-0.3701
-0.3521	-0.3521
-0.9904	-0.8423
1.2505	1.1558
-0.0426	-0.0655
-0.0989	-0.1824
1.1618	1.1401
0.8637	0.8491
-0.4689	-0.4689
-1.1823	-1.3312
-1.4719	-1.4942
-0.7312	-0.8972
0.2657	0.2657
0.2147	0.2889
0.6908	0.5215
0.6558	0.6558

-0.7905	-0.8882
-0.5053	-0.3189
-0.1646	-0.3456
0.0046	0.0046
2.2924	2.3016
-0.3898	-0.4399
0.4266	0.4266
0.4110	0.3458
0.7494	0.7643
1.5156	1.4916
0.2676	0.1243
0.0193	-0.0320
Correlation =.99639	

Grade 3 Social Studies	
1000 Random	Population
0.4836	0.5435
-0.0121	0.0328
0.2404	0.2947
0.4025	0.4025
-1.6166	-1.4123
-1.0125	-0.8347
-0.6485	-0.5630
0.0089	0.0089
-0.9113	-0.8821
-0.1624	-0.2003
-0.1414	-0.1414
0.1045	0.0748
-0.3946	-0.4417
-0.8998	-0.9116
0.7570	0.7570
-0.4238	-0.3103
0.8935	0.7054
-0.0091	-0.0091
0.0739	0.1567
-0.9309	-0.9497
0.1366	0.2077
-0.5480	-0.5480
0.7726	0.8475
1.1771	1.0552
0.0335	0.0335
-1.8588	-1.6906
-0.1802	-0.1863
0.3201	0.3387
1.1123	1.0014
0.9060	0.8869
0.1005	0.1800
0.5838	0.6245
-1.1162	-1.0357
-0.1503	-0.1503
1.0423	0.9698
1.0842	1.0155
-0.0473	-0.0939
-0.0625	-0.0625
-0.3606	-0.3402
-0.2755	-0.2462
0.3284	0.3284
0.7505	0.7572
0.8659	0.9074
0.7285	0.6423
-0.3212	-0.3212
-1.3448	-1.3259
0.1106	0.1016
-0.0655	-0.0655

-0.5494	-0.4291
-0.6941	-0.5861
0.4971	0.4040
0.1759	0.1759
-0.7042	-0.7404
0.4491	0.4776
-0.0790	-0.0790
-0.3468	-0.3863
-1.4948	-1.3742
-0.3460	-0.2373
-0.8465	-0.9616
-0.6089	-0.6802
	Correlation=.99556

Grade 5 ELA	
1000 Random	Population
0.4787	0.4693
0.7810	0.7300
-0.7209	-0.6712
-0.9890	-1.1305
-2.4102	-2.5275
-1.5308	-1.4989
-0.2127	-0.2127
0.0626	0.1389
1.0525	0.9410
-0.0149	-0.0149
0.3174	0.4078
-1.2497	-1.2497
-1.1231	-1.1735
0.9660	0.9678
1.0595	1.0595
-0.5651	-0.6464
-0.8820	-0.8820
0.5293	0.5600
-0.5871	-0.5293
-0.0400	-0.1945
1.0326	0.9794
0.4593	0.3851
0.8567	0.9581
1.5866	1.5686
0.7884	0.8401
1.3198	1.3643
-0.5143	-0.5232
-0.6740	-0.5352
0.5026	0.4900
0.0925	-0.0791
0.6387	0.6384
0.6640	0.6640
-0.7281	-0.6908
1.2328	1.1276
0.7456	0.7456
1.2440	1.2867
-0.3244	-0.3244
0.2659	0.1192
0.8464	0.8807
0.1202	0.1202
0.3032	0.2097
0.3542	0.3542
0.7590	0.7351
0.1618	0.2176
0.9384	0.9384
1.5241	1.4925
-0.2253	-0.1411
-0.4048	-0.3579

-0.4791	-0.5169
-0.3700	-0.3723
Correlation=.99678	

Grade 5 Math	
1000 Random	Population
-0.1528	-0.0212
-0.2763	-0.3023
-0.5225	-0.5271
-0.9749	-0.9749
2.0803	2.0168
-0.3429	-0.3990
-1.1621	-1.0498
-0.5006	-0.5006
0.2214	0.1955
-0.0075	-0.0205
-0.4336	-0.4336
0.5697	0.6984
-0.1820	-0.0782
0.2493	0.1532
-0.0491	-0.0491
1.1933	1.3679
0.6027	0.6693
0.3815	0.3815
1.0842	1.1754
-0.5635	-0.6070
-0.7489	-0.7172
0.5562	0.5562
-0.1118	-0.0374
-0.6937	-0.6943
-0.5080	-0.5080
0.3460	0.4197
0.5793	0.6506
-1.6900	-1.7767
0.8027	0.7418
1.0028	0.8829
-0.3016	-0.3925
-0.5674	-0.7227
-2.0743	-1.9760
0.1860	0.1860
-0.9253	-0.8323
1.0252	0.9379
-0.4609	-0.4701
0.4894	0.4894
0.9783	1.0948
-0.3483	-0.3487
-1.1403	-1.1403
-0.0181	0.1370
-0.0898	-0.1962
-1.2551	-1.2047
-0.4704	-0.4704
-1.3719	-1.3684
-0.7922	-0.7339
0.4108	0.4108

-0.1060	-0.0598
-0.2675	-0.2812
0.6762	0.7765
0.2861	0.2861
-1.0326	-0.8634
-1.2054	-1.1528
0.3872	0.3872
-0.1512	-0.2591
0.8154	0.8275
1.1015	0.9997
0.4510	0.5907
-0.6442	-0.5548
	Correlation=.99525

Grade 5 Social Studies	
1000 Random	Population
-1.5467	-1.4684
-0.8096	-0.7996
-0.4990	-0.5641
0.6745	0.6745
-0.7897	-0.7267
-0.2804	-0.2387
-0.7481	-0.7233
-0.2149	-0.2149
-1.2893	-1.3486
-0.0863	-0.1361
0.4156	0.4156
1.5695	1.5556
0.2684	0.3446
0.1736	0.2317
0.7886	0.7886
0.4719	0.5112
-2.5454	-2.6093
0.1123	0.1123
-0.1725	-0.1185
0.6355	0.7216
0.0587	0.1402
-0.8865	-0.8865
0.5826	0.7311
1.5254	1.3977
0.4337	0.4337
1.1429	1.2114
-0.0580	-0.1377
1.1652	1.1651
0.3261	0.3939
-0.3193	-0.2578
1.6757	1.6858
0.7484	0.6557
1.2271	1.4059
-0.0443	-0.0443
1.4861	1.4389
0.2542	0.2723
1.4386	1.4402
-0.2506	-0.2506
-0.0615	0.0031
1.2596	1.2197
0.4924	0.4924
1.1163	1.1665
0.4383	0.4558
1.0395	1.0261
-0.7970	-0.7970
0.4199	0.3784
-0.2426	-0.2895
-0.2998	-0.2998

0.7898	0.8640
-1.2486	-1.2254
-0.0392	0.0860
0.1534	0.1534
0.4539	0.3688
0.6244	0.5117
-1.0621	-1.0621
0.0879	0.0401
-0.2426	-0.2083
-0.3024	-0.2929
0.7448	0.6598
-1.7895	-1.7858
Correlation=.99760	

Grade 7 ELA	
1000 Random	Population
-0.3023	-0.2541
-2.3221	-2.1845
-0.6828	-0.6632
-0.5272	-0.4291
-2.1307	-2.1335
-1.8651	-1.8110
-0.5566	-0.5566
1.2109	1.1409
0.6388	0.6139
0.7278	0.7278
-0.9907	-1.0763
1.0452	1.0452
0.1084	0.1338
-0.3154	-0.2499
-0.7757	-0.7757
-0.4384	-0.5074
0.8545	0.8545
-0.7197	-0.7200
-1.0117	-0.9875
0.9029	0.9029
-0.9820	-0.9751
0.1864	0.0700
-0.9435	-0.8982
1.1908	1.1081
-0.0345	0.0396
-0.1187	-0.0201
0.3259	0.3127
-0.5201	-0.5695
0.6899	0.6412
1.6584	1.6243
-0.6067	-0.7523
0.6477	0.6477
-0.6890	-0.6936
-0.3108	-0.1848
-0.3628	-0.3628
-0.6864	-0.7133
0.1345	0.1345
0.6201	0.6662
1.2937	1.2438
-0.2009	-0.2009
0.1374	0.2036
-0.4600	-0.4600
0.4307	0.4186
-0.0436	-0.0594
0.6319	0.6319
0.8381	0.7988

0.2175	0.2713
0.6288	0.6902
1.1690	1.2268
0.6224	0.5758
Correlation=.99795	

Grade 7 Math	
1000 Random	Population
0.6216	0.6360
-0.4144	-0.4537
-0.7177	-0.7145
0.2424	0.2424
-0.4331	-0.4068
-0.7733	-0.8405
0.6923	0.6605
-0.4281	-0.4281
1.4647	1.4844
-0.1453	-0.1391
-0.2575	-0.2575
-1.4616	-1.4017
0.6214	0.7300
-0.2540	-0.2127
-0.3464	-0.3464
-1.0596	-0.9906
-1.9846	-2.0717
0.0257	0.0257
0.9111	0.8800
-0.3591	-0.4631
-0.9206	-0.8309
-0.8260	-0.8260
0.8792	0.8504
-0.3306	-0.1991
-0.1291	-0.1291
-0.6056	-0.7588
0.5151	0.5284
-0.7860	-0.7974
-0.3283	-0.3020
-0.1401	-0.1882
1.0522	0.9687
-1.2255	-1.2258
0.3696	0.2537
-1.0170	-1.0170
-0.8401	-0.7901
-0.2890	-0.3401
0.9567	0.8598
0.1007	0.1007
0.2859	0.1934
0.1211	0.1382
-0.3377	-0.3377
-1.4411	-1.3223
-0.4444	-0.2798
1.7564	1.8780
0.4820	0.4820
0.1438	0.1613

0.1493	0.2009
-0.0533	-0.0533
0.8917	0.9947
-0.6898	-0.5252
0.8003	0.8204
0.7921	0.7921
0.9805	0.9998
-0.4078	-0.4097
0.5350	0.5350
-0.5140	-0.4215
-0.0614	-0.0036
2.0956	2.1512
2.1702	1.9609
1.4718	1.4690
Correlation=.99662	

Grade 7 Social Studies	
1000 Random	Population
-0.1187	-0.1216
0.3013	0.2429
-0.4346	-0.3997
0.2097	0.2097
0.5546	0.5296
0.0552	0.2421
0.9217	0.9062
-0.1096	-0.1096
0.2449	0.2347
-0.5592	-0.5497
0.6519	0.6519
0.8441	0.7945
-0.1745	-0.1521
-1.5303	-1.4269
-0.3410	-0.3410
-0.0603	-0.0161
0.4694	0.4183
0.0301	0.0301
-0.2889	-0.3962
-0.1013	-0.1627
0.6178	0.5214
-1.5582	-1.5582
0.3822	0.4428
-0.2248	-0.2549
-0.2566	-0.2566
0.0268	-0.0260
0.6029	0.4874
0.5047	0.5343
-1.2743	-1.4064
-0.6324	-0.6497
-0.6881	-0.5672
-0.3764	-0.2990
0.2770	0.2381
-0.6379	-0.6379
0.3930	0.5122
0.0038	0.1022
0.2268	0.1324
-0.8431	-0.8431
-0.0764	-0.0980
-0.7502	-0.6439
0.0163	0.0163
0.9809	0.8977
0.2094	0.2211
0.1914	0.1565
0.6363	0.6363
0.6563	0.7230
0.6671	0.6394
0.7066	0.7066

0.5134	0.4968
-0.9766	-0.8794
-0.4184	-0.4543
0.3276	0.3276
-0.6148	-0.6864
0.1627	0.1044
0.3333	0.4582
0.5930	0.4683
-0.7933	-0.8351
-0.9280	-0.9197
0.3598	0.2964
0.5764	0.7027
	Correlation=.99353

APPENDIX F

RESULTS OF DIFFERNTIAL ITEM AND TEST FUNCTIONING ANALYSES

Grade 3 ELA

Item	Mean(d)	SD(d)	CDIF	NCDIF	Sig.
1	-0.0060	0.0090	-0.0012	0.0001	ns
2	0.1590	0.1320	-0.0191	0.0429	0.001
3	0.0000	0.0000	0.0000	0.0000	ns
4	-0.0010	0.0010	0.0006	0.0000	ns
5	0.0450	0.0660	0.0082	0.0063	0.001
6	0.0000	0.0000	0.0000	0.0000	ns
7	0.0830	0.0540	-0.0259	0.0098	0.001
8	0.0300	0.0260	-0.0030	0.0016	0.05
9	0.0000	0.0000	0.0000	0.0000	ns
10	-0.0770	0.0420	0.0355	0.0078	0.001
11	-0.1080	0.0650	0.0394	0.0159	0.001
12	0.0000	0.0000	0.0000	0.0000	ns
13	-0.0310	0.0370	-0.0031	0.0023	0.01
14	0.0120	0.0150	0.0013	0.0004	ns
15	0.0000	0.0000	0.0000	0.0000	ns
16	-0.0460	0.0260	0.0197	0.0028	0.01
17	-0.1220	0.0590	0.0703	0.0183	0.001
18	0.0000	0.0000	0.0000	0.0000	ns
19	-0.0600	0.0600	0.0006	0.0072	0.001
20	-0.0870	0.0520	0.0330	0.0102	0.001
21	0.0090	0.0080	-0.0006	0.0002	ns
22	0.0080	0.0040	-0.0048	0.0001	ns
23	-0.0370	0.0220	0.0147	0.0019	0.05
24	0.0280	0.0250	-0.0024	0.0014	0.05
25	0.0140	0.0070	-0.0090	0.0003	ns
26	-0.0980	0.0540	0.0431	0.0124	0.001
27	0.0200	0.0220	0.0010	0.0009	ns
28	-0.0240	0.0160	0.0064	0.0008	ns
29	-0.0600	0.0300	0.0320	0.0045	0.001
30	0.0600	0.0380	-0.0195	0.0050	0.001
31	-0.1880	0.0920	0.1233	0.0436	0.001
32	0.0000	0.0000	0.0000	0.0000	ns
33	-0.0340	0.0220	0.0236	0.0017	0.05
34	-0.0140	0.0110	0.0018	0.0003	ns
35	0.0000	0.0000	0.0000	0.0000	ns

36	-0.0690	0.0410	0.0474	0.0064	0.001
37	0.0000	0.0000	0.0000	0.0000	ns
38	-0.0130	0.0060	0.0082	0.0002	ns
39	0.0780	0.0570	-0.0174	0.0093	0.001
40	0.0000	0.0000	0.0000	0.0000	ns
41	0.0690	0.0440	-0.0234	0.0067	0.001
42	0.0000	0.0000	0.0000	0.0000	ns
43	-0.0050	0.0030	0.0019	0.0000	ns
44	0.0350	0.0280	-0.0057	0.0020	0.05
45	0.0000	0.0000	0.0000	0.0000	ns
46	0.0520	0.0360	-0.0137	0.0040	0.01
47	-0.1240	0.0610	0.0810	0.0192	0.001
48	0.0040	0.0020	-0.0020	0.0000	ns
49	-0.0170	0.0080	0.0103	0.0004	ns
50	0.0830	0.0570	-0.0220	0.0101	0.001
			Differential test functioning (DTF):		.43054

Grade 3 Math

Item	Mean(d)	SD(d)	CDIF	NCDIF	Sig.
1	0.0000	0.0000	0.0000	0.0000	ns
2	0.0700	0.1020	-0.0058	0.0152	0.001
3	0.0160	0.0180	-0.0050	0.0006	ns
4	0.0360	0.0240	-0.0248	0.0019	0.05
5	0.0610	0.0350	-0.0453	0.0049	0.001
6	0.0340	0.0230	-0.0218	0.0017	0.05
7	0.0020	0.0030	-0.0004	0.0000	ns
8	0.0000	0.0000	0.0000	0.0000	ns
9	0.0050	0.0110	0.0008	0.0001	ns
10	0.0930	0.0510	-0.0698	0.0113	0.001
11	0.0660	0.0900	-0.0085	0.0124	0.001
12	0.0000	0.0000	0.0000	0.0000	ns
13	0.0040	0.0030	-0.0029	0.0000	ns
14	0.0450	0.0730	0.0003	0.0073	0.001
15	0.0010	0.0010	-0.0004	0.0000	ns
16	0.0000	0.0000	0.0000	0.0000	ns
17	0.0180	0.0200	-0.0052	0.0007	ns
18	0.1120	0.1360	-0.0253	0.0312	0.001
19	-0.0090	0.0060	0.0069	0.0001	ns
20	0.0000	0.0000	0.0000	0.0000	ns
21	-0.2140	0.1320	0.1457	0.0633	0.001
22	-0.1030	0.0920	0.0498	0.0190	0.001
23	-0.0370	0.0640	-0.0016	0.0054	0.001
24	0.0000	0.0000	0.0000	0.0000	ns
25	-0.0780	0.0420	0.0582	0.0078	0.001
26	0.0840	0.0560	-0.0566	0.0102	0.001
27	0.0340	0.0200	-0.0248	0.0016	ns
28	0.0480	0.0270	-0.0361	0.0030	0.01
29	-0.0510	0.0360	0.0331	0.0039	0.001
30	-0.0630	0.0370	0.0443	0.0053	0.001
31	-0.1150	0.0620	0.0858	0.0169	0.001
32	-0.0590	0.0380	0.0390	0.0049	0.001
33	-0.0170	0.0170	0.0069	0.0006	ns
34	0.0000	0.0000	0.0000	0.0000	ns
35	-0.0060	0.0080	0.0011	0.0001	ns
36	-0.0160	0.0090	0.0119	0.0003	ns
37	0.0560	0.0420	-0.0335	0.0048	0.001
38	0.0200	0.0160	-0.0111	0.0007	ns
39	-0.0560	0.0310	0.0424	0.0042	0.001
40	-0.0430	0.0250	0.0316	0.0024	0.05
41	0.0000	0.0000	0.0000	0.0000	ns
42	0.0260	0.0340	-0.0044	0.0019	0.05
43	0.0430	0.0600	-0.0050	0.0055	0.001
44	0.0090	0.0100	-0.0027	0.0002	ns
45	0.0000	0.0000	0.0000	0.0000	ns
46	-0.0790	0.0620	0.0450	0.0100	0.001

47	-0.0680	0.0420	0.0478	0.0063	0.001
48	0.0000	0.0000	0.0000	0.0000	ns
49	-0.0630	0.0870	0.0082	0.0115	0.001
50	0.1070	0.0910	-0.0546	0.0198	0.001
51	-0.0320	0.0290	0.0154	0.0019	0.05
52	0.0000	0.0000	0.0000	0.0000	ns
53	-0.0180	0.0130	0.0113	0.0005	ns
54	0.0010	0.0010	-0.0005	0.0000	ns
55	0.0000	0.0000	0.0000	0.0000	ns
56	-0.0970	0.0700	0.0605	0.0144	0.001
57	-0.1440	0.0920	0.0981	0.0293	0.001
58	-0.0260	0.0140	0.0192	0.0009	ns
59	-0.0910	0.0700	0.0528	0.0133	0.001
60	-0.0430	0.0360	0.0232	0.0032	0.01
			Differential test functioning (DTF):		.49281

Grade 3 Social Studies

Item	Mean(d)	SD(d)	CDIF	NCDIF	Sig.
1	0.0940	0.0440	0.1494	0.0108	0.001
2	0.1750	0.0800	0.3076	0.0369	0.001
3	-0.0130	0.0060	-0.0232	0.0002	ns
4	0.0000	0.0000	0.0000	0.0000	ns
5	0.1270	0.1230	0.2739	0.0314	0.001
6	0.1180	0.0920	0.2510	0.0225	0.001
7	0.0930	0.0630	0.1929	0.0127	0.001
8	0.0000	0.0000	0.0000	0.0000	ns
9	0.1360	0.0980	0.2856	0.0283	0.001
10	0.1210	0.0610	0.2271	0.0183	0.001
11	0.0000	0.0000	0.0000	0.0000	ns
12	0.0040	0.0020	0.0070	0.0000	ns
13	0.0470	0.0290	0.0949	0.0030	0.01
14	0.0800	0.0620	0.1690	0.0102	0.001
15	0.0000	0.0000	0.0000	0.0000	ns
16	0.0170	0.0110	0.0348	0.0004	ns
17	-0.0080	0.0040	-0.0112	0.0001	ns
18	0.0000	0.0000	0.0000	0.0000	ns
19	0.0680	0.0330	0.1249	0.0058	0.001
20	0.0870	0.0680	0.1838	0.0121	0.001
21	0.0370	0.0180	0.0669	0.0017	0.05
22	0.0000	0.0000	0.0000	0.0000	ns
23	0.0460	0.0230	0.0685	0.0027	0.05
24	-0.0360	0.0190	-0.0486	0.0017	0.05
25	0.0000	0.0000	0.0000	0.0000	ns
26	0.1510	0.1510	0.3261	0.0456	0.001
27	-0.0100	0.0060	-0.0207	0.0002	ns
28	-0.1090	0.0550	-0.2049	0.0150	0.001
29	0.0190	0.0100	0.0251	0.0005	ns
30	-0.0240	0.0120	-0.0349	0.0007	ns
31	0.0590	0.0290	0.1074	0.0043	0.001
32	-0.0040	0.0020	-0.0063	0.0000	ns
33	0.0310	0.0290	0.0678	0.0019	0.05
34	0.0000	0.0000	0.0000	0.0000	ns
35	-0.0160	0.0080	-0.0220	0.0003	ns
36	-0.0200	0.0100	-0.0273	0.0005	ns
37	0.0200	0.0110	0.0389	0.0005	ns
38	0.0000	0.0000	0.0000	0.0000	ns
39	0.0220	0.0140	0.0457	0.0007	ns
40	0.0750	0.0430	0.1485	0.0075	0.001
41	0.0000	0.0000	0.0000	0.0000	ns
42	0.0830	0.0410	0.1195	0.0085	0.001
43	-0.0190	0.0090	-0.0285	0.0005	ns
44	0.0240	0.0120	0.0368	0.0007	ns
45	0.0000	0.0000	0.0000	0.0000	ns
46	-0.0310	0.0360	-0.0667	0.0023	0.01

47	-0.0210	0.0110	-0.0396	0.0006	ns
48	0.0000	0.0000	0.0000	0.0000	ns
49	0.0350	0.0240	0.0729	0.0018	0.05
50	-0.0070	0.0060	-0.0153	0.0001	ns
51	-0.0650	0.0300	-0.1130	0.0051	0.001
52	0.0000	0.0000	0.0000	0.0000	ns
53	0.0450	0.0340	0.0953	0.0032	0.01
54	0.1280	0.0590	0.2007	0.0199	0.001
55	0.0000	0.0000	0.0000	0.0000	ns
56	0.0290	0.0180	0.0585	0.0012	ns
57	-0.0020	0.0030	-0.0050	0.0000	ns
58	0.0160	0.0100	0.0327	0.0004	ns
59	0.0630	0.0490	0.1338	0.0064	0.001
60	-0.1300	0.1200	-0.2799	0.0313	0.001
				Differential test functioning (DTF):	2.9996

Grade 5 ELA

Item	Mean(d)	SD(d)	CDIF	NCDIF	Sig.
1	0.0160	0.0090	-0.0108	0.0003	ns
2	-0.0750	0.0390	0.0505	0.0071	0.001
3	0.1180	0.0950	-0.0688	0.0229	0.001
4	-0.0080	0.0100	0.0039	0.0002	ns
5	0.0990	0.1310	-0.0407	0.0271	0.001
6	0.0120	0.0170	-0.0048	0.0005	ns
7	0.0000	0.0000	0.0000	0.0000	ns
8	0.1530	0.0840	-0.1020	0.0306	0.001
9	-0.0370	0.0180	0.0244	0.0017	0.05
10	0.0000	0.0000	0.0000	0.0000	ns
11	0.0050	0.0030	-0.0032	0.0000	ns
12	0.0000	0.0000	0.0000	0.0000	ns
13	-0.0160	0.0210	0.0070	0.0007	ns
14	-0.0280	0.0140	0.0184	0.0010	ns
15	0.0000	0.0000	0.0000	0.0000	ns
16	-0.0130	0.0120	0.0071	0.0003	ns
17	0.0000	0.0000	0.0000	0.0000	ns
18	-0.0400	0.0220	0.0271	0.0021	0.05
19	0.0360	0.0320	-0.0206	0.0023	0.01
20	-0.0260	0.0180	0.0162	0.0010	ns
21	0.0590	0.0290	-0.0380	0.0044	0.001
22	-0.0210	0.0120	0.0142	0.0006	ns
23	-0.0350	0.0170	0.0232	0.0015	ns
24	-0.0280	0.0150	0.0164	0.0010	ns
25	-0.0530	0.0270	0.0359	0.0036	0.001
26	-0.0460	0.0230	0.0288	0.0026	0.05
27	-0.0160	0.0150	0.0088	0.0005	ns
28	0.0370	0.0340	-0.0207	0.0026	0.01
29	0.0190	0.0100	-0.0128	0.0005	ns
30	-0.0130	0.0090	0.0084	0.0002	ns
31	-0.0200	0.0110	0.0138	0.0005	ns
32	0.0000	0.0000	0.0000	0.0000	ns
33	-0.0070	0.0070	0.0034	0.0001	ns
34	-0.0270	0.0130	0.0172	0.0009	ns
35	0.0000	0.0000	0.0000	0.0000	ns
36	-0.0300	0.0150	0.0191	0.0011	ns
37	0.0000	0.0000	0.0000	0.0000	ns
38	-0.0420	0.0260	0.0273	0.0024	0.05
39	-0.1040	0.0530	0.0696	0.0135	0.001
40	0.0000	0.0000	0.0000	0.0000	ns
41	-0.0510	0.0310	0.0334	0.0036	0.001
42	0.0000	0.0000	0.0000	0.0000	ns
43	-0.1800	0.0980	0.1198	0.0420	0.001
44	-0.0340	0.0220	0.0220	0.0016	0.05
45	0.0000	0.0000	0.0000	0.0000	ns
46	0.0250	0.0140	-0.0144	0.0008	ns

47	-0.0200	0.0160	0.0121	0.0007	ns
48	-0.0700	0.0660	0.0374	0.0092	0.001
49	-0.0390	0.0360	0.0211	0.0028	0.01
50	-0.0250	0.0210	0.0143	0.0011	ns
			Differential test functioning (DTF):		.3637

Grade 5 Math

Item	Mean(d)	SD(d)	CDIF	NCDIF	Sig.
1	0.1420	0.0730	-0.0032	0.0255	0.001
2	0.0170	0.0110	0.0008	0.0004	ns
3	0.0860	0.0560	0.0051	0.0106	0.001
4	0.0000	0.0000	0.0000	0.0000	ns
5	-0.2310	0.1340	0.0559	0.0714	0.001
6	0.0300	0.0190	0.0016	0.0013	ns
7	0.0270	0.0260	0.0042	0.0014	0.05
8	0.0000	0.0000	0.0000	0.0000	ns
9	0.0860	0.0410	-0.0060	0.0090	0.001
10	-0.0250	0.0140	-0.0004	0.0008	ns
11	0.0000	0.0000	0.0000	0.0000	ns
12	0.0950	0.0450	-0.0130	0.0111	0.001
13	0.2210	0.1060	-0.0102	0.0601	0.001
14	-0.0800	0.0420	0.0011	0.0083	0.001
15	0.0000	0.0000	0.0000	0.0000	ns
16	0.0170	0.0090	-0.0036	0.0004	ns
17	0.0280	0.0130	-0.0035	0.0010	ns
18	0.0000	0.0000	0.0000	0.0000	ns
19	0.0320	0.0170	-0.0066	0.0013	ns
20	0.0690	0.0470	0.0050	0.0069	0.001
21	0.1250	0.0850	0.0095	0.0230	0.001
22	0.0000	0.0000	0.0000	0.0000	ns
23	0.0660	0.0360	-0.0001	0.0057	0.001
24	-0.0290	0.0250	-0.0037	0.0015	0.05
25	0.0000	0.0000	0.0000	0.0000	ns
26	-0.0330	0.0160	0.0017	0.0014	ns
27	-0.0520	0.0250	0.0047	0.0033	0.05
28	-0.0260	0.0380	-0.0053	0.0021	0.01
29	-0.1670	0.0770	0.0153	0.0339	0.001
30	-0.1970	0.0900	0.0232	0.0468	0.001
31	0.0640	0.0380	0.0021	0.0055	0.001
32	-0.0020	0.0010	-0.0002	0.0000	ns
33	0.0540	0.0720	0.0107	0.0082	0.001
34	0.0000	0.0000	0.0000	0.0000	ns
35	0.0300	0.0260	0.0040	0.0016	0.05
36	0.0220	0.0110	-0.0042	0.0006	ns
37	0.0450	0.0300	0.0030	0.0029	0.01
38	0.0000	0.0000	0.0000	0.0000	ns
39	-0.0100	0.0050	0.0018	0.0001	ns
40	-0.0510	0.0360	-0.0043	0.0039	0.001
41	0.0000	0.0000	0.0000	0.0000	ns
42	-0.0050	0.0030	-0.0001	0.0000	ns
43	-0.1220	0.0810	-0.0083	0.0215	0.001
44	-0.0320	0.0390	-0.0060	0.0025	0.01
45	0.0000	0.0000	0.0000	0.0000	ns
46	0.0530	0.0550	0.0087	0.0058	0.001

47	-0.0200	0.0180	-0.0027	0.0007	ns
48	0.0000	0.0000	0.0000	0.0000	ns
49	0.0070	0.0040	0.0001	0.0001	ns
50	-0.1000	0.0730	-0.0091	0.0153	0.001
51	-0.0960	0.0450	0.0089	0.0112	0.001
52	0.0000	0.0000	0.0000	0.0000	ns
53	0.0020	0.0020	0.0003	0.0000	ns
54	0.0520	0.0500	0.0079	0.0052	0.001
55	0.0000	0.0000	0.0000	0.0000	ns
56	-0.0800	0.0520	-0.0050	0.0091	0.001
57	0.0040	0.0020	-0.0007	0.0000	ns
58	-0.0750	0.0370	0.0131	0.0070	0.001
59	0.0320	0.0150	-0.0030	0.0012	ns
60	-0.0620	0.0540	-0.0084	0.0068	0.001
			Differential test functioning (DTF):		.0813

Grade 5 Social Studies

Item	Mean(d)	SD(d)	CDIF	NCDIF	Sig.
1	0.7360	0.1610	0.1096	0.5674	0.001
2	-0.0110	0.0080	-0.0044	0.0002	ns
3	0.0590	0.0320	0.0184	0.0045	0.001
4	0.0000	0.0000	0.0000	0.0000	ns
5	0.1800	0.0970	0.0565	0.0418	0.001
6	0.0000	0.0000	0.0000	0.0000	ns
7	0.1140	0.0660	0.0384	0.0175	0.001
8	0.0000	0.0000	0.0000	0.0000	ns
9	0.0750	0.0610	0.0316	0.0094	0.001
10	-0.0340	0.0160	-0.0086	0.0014	ns
11	0.0000	0.0000	0.0000	0.0000	ns
12	-0.0120	0.0080	0.0022	0.0002	ns
13	-0.1620	0.0680	-0.0352	0.0309	0.001
14	-0.0400	0.0160	-0.0073	0.0018	0.05
15	0.0000	0.0000	0.0000	0.0000	ns
16	0.0550	0.0220	0.0024	0.0035	0.01
17	-0.0040	0.0050	-0.0017	0.0000	ns
18	0.0000	0.0000	0.0000	0.0000	ns
19	0.0580	0.0260	0.0135	0.0040	0.01
20	0.0450	0.0190	0.0000	0.0024	0.05
21	0.0500	0.0200	0.0086	0.0029	0.01
22	0.0000	0.0000	0.0000	0.0000	ns
23	0.0730	0.0300	0.0001	0.0063	0.001
24	-0.0220	0.0150	0.0040	0.0007	ns
25	0.0000	0.0000	0.0000	0.0000	ns
26	-0.0470	0.0240	0.0043	0.0028	0.01
27	0.0140	0.0060	0.0031	0.0002	ns
28	-0.0090	0.0050	0.0010	0.0001	ns
29	0.0120	0.0050	0.0014	0.0002	ns
30	0.1280	0.0560	0.0303	0.0196	0.001
31	-0.0780	0.0520	0.0141	0.0088	0.001
32	-0.0500	0.0200	-0.0008	0.0029	0.01
33	-0.0190	0.0100	0.0023	0.0005	ns
34	0.0000	0.0000	0.0000	0.0000	ns
35	-0.0760	0.0460	0.0114	0.0078	0.001
36	0.0140	0.0050	0.0018	0.0002	ns
37	-0.0510	0.0310	0.0078	0.0036	0.01
38	0.0000	0.0000	0.0000	0.0000	ns
39	0.0670	0.0280	0.0133	0.0053	0.001
40	-0.1070	0.0540	0.0095	0.0142	0.001
41	0.0000	0.0000	0.0000	0.0000	ns
42	-0.0320	0.0160	0.0030	0.0013	ns
43	-0.0870	0.0330	-0.0112	0.0086	0.001
44	-0.0720	0.0330	0.0038	0.0064	0.001
45	0.0000	0.0000	0.0000	0.0000	ns
46	-0.0200	0.0080	-0.0020	0.0005	ns
47	-0.0230	0.0120	-0.0068	0.0007	ns

48	0.0000	0.0000	0.0000	0.0000	ns
49	-0.0380	0.0160	0.0001	0.0017	0.05
50	-0.0700	0.0710	-0.0313	0.0098	0.001
51	0.1170	0.0460	0.0195	0.0159	0.001
52	0.0000	0.0000	0.0000	0.0000	ns
53	-0.0420	0.0160	-0.0042	0.0020	0.05
54	-0.1340	0.0500	-0.0131	0.0205	0.001
55	0.0000	0.0000	0.0000	0.0000	ns
56	-0.0330	0.0140	-0.0068	0.0013	ns
57	-0.0410	0.0220	-0.0124	0.0022	0.05
58	-0.0430	0.0230	-0.0133	0.0024	0.05
59	-0.3130	0.1110	-0.0481	0.1100	0.001
60	-0.0330	0.0410	-0.0154	0.0028	0.001
			Differential test functioning (DTF):		.1891

Grade 7 ELA

Item	Mean(d)	SD(d)	CDIF	NCDIF	Sig.
1	0.0390	0.0310	0.0138	0.0024	0.01
2	0.0450	0.0760	0.0361	0.0079	0.001
3	0.0060	0.0060	0.0031	0.0001	ns
4	0.0340	0.0300	0.0144	0.0020	0.05
5	0.0640	0.0970	0.0478	0.0134	0.001
6	0.1400	0.1610	0.0844	0.0455	0.001
7	0.0000	0.0000	0.0000	0.0000	ns
8	-0.1490	0.0780	-0.0174	0.0282	0.001
9	0.0030	0.0010	0.0004	0.0000	ns
10	0.0000	0.0000	0.0000	0.0000	ns
11	0.0230	0.0250	0.0130	0.0012	ns
12	0.0000	0.0000	0.0000	0.0000	ns
13	-0.0360	0.0260	-0.0106	0.0020	0.05
14	0.0340	0.0270	0.0124	0.0019	0.05
15	0.0000	0.0000	0.0000	0.0000	ns
16	-0.0430	0.0410	-0.0208	0.0036	0.01
17	0.0000	0.0000	0.0000	0.0000	ns
18	0.0000	0.0000	-0.0001	0.0000	ns
19	0.0340	0.0380	0.0194	0.0026	0.001
20	0.0000	0.0000	0.0000	0.0000	ns
21	-0.0050	0.0060	-0.0030	0.0001	ns
22	-0.0250	0.0170	-0.0067	0.0009	ns
23	0.0310	0.0340	0.0173	0.0021	0.05
24	-0.0150	0.0080	-0.0014	0.0003	ns
25	-0.0240	0.0180	-0.0078	0.0009	ns
26	0.0240	0.0170	0.0074	0.0009	ns
27	-0.0070	0.0040	-0.0014	0.0001	ns
28	-0.0090	0.0080	-0.0042	0.0002	ns
29	-0.0690	0.0400	-0.0118	0.0063	0.001
30	0.0550	0.0340	0.0033	0.0042	0.001
31	-0.0450	0.0470	-0.0242	0.0042	0.01
32	0.0000	0.0000	0.0000	0.0000	ns
33	0.0240	0.0230	0.0117	0.0011	ns
34	0.0360	0.0280	0.0129	0.0021	0.05
35	0.0000	0.0000	0.0000	0.0000	ns
36	-0.0320	0.0340	-0.0176	0.0022	0.05
37	0.0000	0.0000	0.0000	0.0000	ns
38	-0.0570	0.0330	-0.0103	0.0044	0.001
39	-0.0150	0.0080	-0.0012	0.0003	ns
40	0.0000	0.0000	0.0000	0.0000	ns
41	-0.0270	0.0190	-0.0075	0.0011	ns
42	0.0000	0.0000	0.0000	0.0000	ns
43	0.0000	0.0000	-0.0001	0.0000	ns
44	0.0150	0.0110	0.0046	0.0004	ns
45	0.0000	0.0000	0.0000	0.0000	ns
46	-0.0410	0.0220	-0.0056	0.0022	0.05
47	0.0380	0.0240	0.0085	0.0021	0.05

48	-0.0100	0.0060	-0.0016	0.0001	ns
49	-0.0500	0.0260	-0.0049	0.0032	0.01
50	0.2290	0.1150	0.0233	0.0658	0.001
			Differential test functioning (DTF):		.1755

Item	Mean(d)	SD(d)	CDIF	NCDIF	Sig.
1	0.1110	0.0590	-0.0117	0.0158	0.001
2	0.0530	0.0320	0.0021	0.0039	0.01
3	0.0200	0.0140	0.0014	0.0006	ns
4	0.0000	0.0000	0.0000	0.0000	ns
5	-0.0030	0.0020	-0.0002	0.0000	ns
6	0.0960	0.0640	0.0056	0.0132	0.001
7	-0.0010	0.0010	0.0001	0.0000	ns
8	0.0000	0.0000	0.0000	0.0000	ns
9	-0.0690	0.0450	0.0116	0.0068	0.001
10	-0.1010	0.0610	-0.0042	0.0138	0.001
11	0.0000	0.0000	0.0000	0.0000	ns
12	0.0340	0.0330	0.0035	0.0023	0.05
13	-0.0210	0.0110	0.0014	0.0006	ns
14	0.0030	0.0020	0.0001	0.0000	ns
15	0.0000	0.0000	0.0000	0.0000	ns
16	0.1250	0.0900	0.0088	0.0238	0.001
17	0.0280	0.0340	0.0033	0.0020	0.01
18	0.0000	0.0000	0.0000	0.0000	ns
19	0.0220	0.0120	-0.0027	0.0006	ns
20	0.0690	0.0400	0.0021	0.0063	0.001
21	-0.0280	0.0230	-0.0024	0.0013	ns
22	0.0000	0.0000	0.0000	0.0000	ns
23	-0.0670	0.0350	0.0060	0.0057	0.001
24	0.0380	0.0220	0.0013	0.0020	0.05
25	0.0000	0.0000	0.0000	0.0000	ns
26	-0.0300	0.0210	-0.0021	0.0014	ns
27	-0.0030	0.0010	0.0002	0.0000	ns
28	-0.0550	0.0440	-0.0046	0.0050	0.001
29	0.0020	0.0010	0.0001	0.0000	ns
30	-0.0290	0.0170	-0.0008	0.0012	ns
31	-0.1750	0.0880	0.0145	0.0385	0.001
32	-0.0280	0.0270	-0.0028	0.0015	0.05
33	-0.1140	0.0580	0.0012	0.0164	0.001
34	0.0000	0.0000	0.0000	0.0000	ns
35	-0.0130	0.0100	-0.0010	0.0003	ns
36	-0.0320	0.0190	-0.0014	0.0014	ns
37	-0.1020	0.0530	0.0092	0.0131	0.001
38	0.0000	0.0000	0.0000	0.0000	ns
39	0.0810	0.0400	-0.0039	0.0082	0.001
40	0.0070	0.0040	-0.0001	0.0001	ns
41	0.0000	0.0000	0.0000	0.0000	ns
42	0.0800	0.0720	0.0075	0.0115	0.001
43	0.0930	0.0540	0.0031	0.0117	0.001
44	0.0730	0.0690	-0.0171	0.0101	0.001
45	0.0000	0.0000	0.0000	0.0000	ns
46	0.0530	0.0270	-0.0011	0.0035	0.01
47	-0.0420	0.0220	0.0000	0.0023	0.05

48	0.0000	0.0000	0.0000	0.0000	ns
49	0.0040	0.0020	-0.0004	0.0000	ns
50	-0.0720	0.0560	-0.0058	0.0084	0.001
51	0.0400	0.0220	-0.0044	0.0021	0.05
52	0.0000	0.0000	0.0000	0.0000	ns
53	-0.0620	0.0340	0.0066	0.0050	0.001
54	-0.1050	0.0730	-0.0068	0.0163	0.001
55	0.0000	0.0000	0.0000	0.0000	ns
56	-0.0360	0.0250	-0.0022	0.0019	0.05
57	0.3110	0.1420	-0.0198	0.1166	0.001
58	-0.0780	0.0680	0.0175	0.0108	0.001
59	-0.1180	0.0990	0.0257	0.0237	0.001
60	-0.0170	0.0120	0.0032	0.0004	ns
			Differential test functioning (DTF):		.0406

Grade 7 Social Studies

Item	Mean(d)	SD(d)	CDIF	NCDIF	Sig.
1	0.1130	0.0430	-0.0433	0.0146	0.001
2	0.0690	0.0290	-0.0268	0.0056	0.001
3	0.1440	0.0590	-0.0548	0.0242	0.001
4	0.0000	0.0000	0.0000	0.0000	ns
5	0.0580	0.0270	-0.0221	0.0041	0.01
6	0.0670	0.0260	-0.0258	0.0051	0.01
7	-0.0360	0.0190	0.0137	0.0017	0.05
8	0.0000	0.0000	0.0000	0.0000	ns
9	-0.1120	0.0430	0.0430	0.0144	0.001
10	0.1580	0.0670	-0.0596	0.0294	0.001
11	0.0000	0.0000	0.0000	0.0000	ns
12	-0.1000	0.0460	0.0382	0.0121	0.001
13	0.0680	0.0270	-0.0262	0.0054	0.001
14	0.0210	0.0180	-0.0074	0.0008	ns
15	0.0000	0.0000	0.0000	0.0000	ns
16	0.0670	0.0260	-0.0259	0.0052	0.001
17	0.0140	0.0060	-0.0055	0.0002	ns
18	0.0000	0.0000	0.0000	0.0000	ns
19	-0.2310	0.1280	0.0846	0.0697	0.001
20	0.0530	0.0200	-0.0202	0.0032	0.01
21	-0.1310	0.0530	0.0505	0.0199	0.001
22	0.0000	0.0000	0.0000	0.0000	ns
23	-0.0470	0.0190	0.0181	0.0026	0.05
24	-0.1190	0.0560	0.0445	0.0173	0.001
25	0.0000	0.0000	0.0000	0.0000	ns
26	-0.0280	0.0110	0.0109	0.0009	ns
27	0.0720	0.0360	-0.0275	0.0065	0.001
28	-0.0110	0.0050	0.0042	0.0001	ns
29	-0.0430	0.0360	0.0151	0.0031	0.001
30	-0.0370	0.0210	0.0135	0.0018	0.05
31	0.0530	0.0280	-0.0196	0.0036	0.01
32	0.0380	0.0170	-0.0144	0.0018	ns
33	-0.1690	0.0640	0.0648	0.0328	0.001
34	0.0000	0.0000	0.0000	0.0000	ns
35	-0.0720	0.0280	0.0279	0.0060	0.001
36	0.0450	0.0170	-0.0171	0.0023	0.05
37	-0.0340	0.0130	0.0131	0.0013	ns
38	0.0000	0.0000	0.0000	0.0000	ns
39	-0.0110	0.0040	0.0041	0.0001	ns
40	0.1050	0.0530	-0.0390	0.0139	0.001
41	0.0000	0.0000	0.0000	0.0000	ns
42	-0.0750	0.0390	0.0282	0.0071	0.001
43	0.0610	0.0240	-0.0235	0.0043	0.01
44	-0.0090	0.0030	0.0033	0.0001	ns
45	0.0000	0.0000	0.0000	0.0000	ns
46	-0.0240	0.0110	0.0091	0.0007	ns

47	-0.1480	0.0590	0.0571	0.0255	0.001
48	0.0000	0.0000	0.0000	0.0000	ns
49	-0.0170	0.0070	0.0067	0.0004	ns
50	0.0120	0.0080	-0.0044	0.0002	ns
51	0.1060	0.0440	-0.0403	0.0133	0.001
52	0.0000	0.0000	0.0000	0.0000	ns
53	-0.0480	0.0270	0.0176	0.0030	0.01
54	-0.1280	0.0500	0.0489	0.0188	0.001
55	0.0910	0.0390	-0.0351	0.0098	0.001
56	-0.1680	0.0650	0.0647	0.0324	0.001
57	-0.0250	0.0160	0.0092	0.0009	ns
58	0.0240	0.0150	-0.0088	0.0008	ns
59	-0.0330	0.0130	0.0127	0.0013	ns
60	0.0600	0.0290	-0.0228	0.0044	0.001
			Differential test functioning (DTF):		.1336

APPENDIX G

RESULTS OF DIFFERNTIAL ITEM AND TEST FUNCTIONING ANALYSES FOR LINK SETS

Grade 3 ELA

Item	Mean(d)	SD(d)	CDIF	NCDIF	Sig.
1	-0.03500	0.03100	-0.00043	0.00218	0.05
2	-0.01200	0.01000	-0.00003	0.00025	ns
3	0.03300	0.02800	0.00025	0.00188	0.05
4	0.02400	0.01700	-0.00015	0.00085	ns
5	-0.01300	0.01100	-0.00011	0.00028	ns
6	-0.01000	0.00800	-0.00009	0.00016	ns
7	-0.01200	0.00900	0.00009	0.00022	ns
8	0.01700	0.01800	0.00042	0.00063	ns
9	0.05300	0.04300	0.00030	0.00463	0.001
10	0.00600	0.00400	-0.00002	0.00005	ns
11	-0.02100	0.01300	0.00050	0.00059	ns
12	-0.04100	0.02600	0.00102	0.00240	0.05
				Differential test functioning (DTF):	0.00175

Grade 3 Math

Item	Mean(d)	SD(d)	CDIF	NCDIF	Sig.
1	-0.03500	0.03100	-0.00043	0.00218	0.05
2	-0.01200	0.01000	-0.00003	0.00025	ns
3	0.03300	0.02800	0.00025	0.00188	0.05
4	0.02400	0.01700	-0.00015	0.00085	ns
5	-0.01300	0.01100	-0.00011	0.00028	ns
6	-0.01000	0.00800	-0.00009	0.00016	ns
7	-0.01200	0.00900	0.00009	0.00022	ns
8	0.01700	0.01800	0.00042	0.00063	ns
9	0.05300	0.04300	0.00030	0.00463	0.001
10	0.00600	0.00400	-0.00002	0.00005	ns
11	-0.02100	0.01300	0.00050	0.00059	ns
12	-0.04100	0.02600	0.00102	0.00240	0.05
				Differential test functioning (DTF):	0.00175

Grade 3 Social Studies

Item	Mean(d)	SD(d)	CDIF	NCDIF	Sig.
1	0.00500	0.00300	-0.00012	0.00004	ns
2	-0.09200	0.05400	0.00221	0.01134	0.001
3	-0.09800	0.06200	0.00279	0.01341	0.001
4	0.03700	0.02200	-0.00032	0.00184	0.05
5	0.01800	0.01400	-0.00056	0.00052	ns
6	-0.01000	0.00900	0.00029	0.00017	ns
7	-0.01500	0.00900	0.00041	0.00032	ns
8	-0.03300	0.02300	0.00102	0.00163	ns
9	0.01300	0.00800	-0.00036	0.00024	ns
10	0.05700	0.03300	-0.00080	0.00437	0.01
11	-0.01000	0.00700	0.00031	0.00015	ns
12	0.02200	0.01500	-0.00067	0.00071	ns
13	0.06900	0.04900	-0.00212	0.00705	0.001
14	0.02800	0.01900	-0.00085	0.00113	ns
			Differential test functioning (DTF):		0.00124

Grade 5 ELA

Item	Mean(d)	SD(d)	CDIF	NCDIF	Sig.
1	0.02200	0.02200	0.00092	0.00096	ns
2	-0.01800	0.01400	-0.00053	0.00050	ns
3	-0.01300	0.01800	-0.00060	0.00049	ns
4	-0.01000	0.00600	0.00010	0.00013	ns
5	-0.00700	0.00800	-0.00030	0.00011	ns
6	0.00200	0.00100	-0.00001	0.00000	ns
7	-0.03500	0.02200	-0.00019	0.00174	0.05
8	0.04400	0.03900	0.00166	0.00351	0.01
9	-0.03200	0.02100	-0.00051	0.00145	ns
10	0.07300	0.05000	0.00128	0.00775	0.001
11	-0.02200	0.01400	-0.00003	0.00065	ns
			Differential test functioning (DTF):		0.00178

Grade 5 Math

Item	Mean(d)	SD(d)	CDIF	NCDIF	Sig.
1	0.01100	0.01100	-0.00002	0.00024	ns
2	-0.06500	0.05200	-0.00074	0.00690	0.001
3	-0.03000	0.02600	-0.00026	0.00155	0.05
4	0.03500	0.02300	0.00079	0.00177	0.05
5	-0.05400	0.03400	-0.00130	0.00411	0.01
6	0.01300	0.00800	0.00037	0.00024	ns
7	0.02000	0.01700	0.00019	0.00068	ns
8	0.06600	0.04200	0.00157	0.00619	0.001
9	0.10000	0.06100	0.00252	0.01368	0.001
10	-0.02000	0.02100	0.00006	0.00085	ns
11	0.02900	0.02600	0.00016	0.00152	0.05
12	-0.07800	0.04700	-0.00207	0.00827	0.001
13	-0.00500	0.00300	-0.00012	0.00003	ns
14	-0.00800	0.00500	-0.00021	0.00008	ns
			Differential test functioning (DTF):		0.00094

Grade 5 Social Studies

Item	Mean(d)	SD(d)	CDIF	NCDIF	Sig.
1	0.01400	0.00800	-0.00076	0.00027	ns
2	-0.01800	0.01100	0.00106	0.00044	ns
3	-0.04000	0.02200	0.00215	0.00210	0.05
4	0.06900	0.04100	-0.00312	0.00648	0.001
5	0.09500	0.05300	-0.00571	0.01187	0.001
6	0.01500	0.01400	-0.00065	0.00042	ns
7	-0.01400	0.00800	0.00073	0.00024	ns
8	-0.10200	0.05600	0.00608	0.01351	0.001
9	-0.01200	0.00800	0.00068	0.00020	ns
10	-0.05600	0.03000	0.00326	0.00411	0.01
11	-0.02300	0.01900	0.00109	0.00088	ns
12	0.02400	0.01300	-0.00142	0.00074	ns
13	-0.02700	0.01500	0.00162	0.00095	ns
14	0.03800	0.03400	-0.00152	0.00261	0.01
			Differential test functioning (DTF):		0.00349

Grade 7 ELA

Item	Mean(d)	SD(d)	CDIF	NCDIF	Sig.
1	-0.01900	0.01900	-0.00004	0.00071	ns
2	-0.01700	0.01000	0.00163	0.00040	ns
3	0.05900	0.03500	-0.00604	0.00471	0.01
4	0.02000	0.02300	0.00056	0.00094	ns
5	-0.05400	0.03200	0.00540	0.00400	0.001
6	-0.07500	0.04400	0.00717	0.00752	0.001
7	-0.02000	0.01300	0.00165	0.00057	ns
8	0.03600	0.03200	-0.00044	0.00232	0.05
9	0.01400	0.01000	-0.00093	0.00031	ns
10	-0.00300	0.00300	0.00004	0.00002	ns
11	0.03900	0.03600	-0.00031	0.00287	0.01
12	-0.03300	0.02000	0.00288	0.00151	ns
				Differential test functioning (DTF):	0.01157

Grade 7 Math

Item	Mean(d)	SD(d)	CDIF	NCDIF	Sig.
1	-0.02500	0.01500	0.00167	0.00085	ns
2	-0.04700	0.03200	0.00138	0.00323	0.01
3	-0.02600	0.01700	0.00108	0.00097	ns
4	-0.00600	0.00400	0.00019	0.00005	ns
5	0.02900	0.01700	-0.00146	0.00110	ns
6	0.07400	0.05900	-0.00112	0.00897	0.001
7	0.00700	0.00400	-0.00035	0.00007	ns
8	0.02800	0.02500	0.00000	0.00143	ns
9	-0.02200	0.01300	0.00116	0.00066	ns
10	-0.00200	0.00100	0.00007	0.00001	ns
11	-0.05700	0.03200	0.00349	0.00423	0.001
12	-0.01400	0.00800	0.00062	0.00026	ns
13	0.00000	0.00000	0.00002	0.00000	ns
14	0.02600	0.01500	-0.00173	0.00092	ns
				Differential test functioning (DTF):	0.00501

Grade 7 Social Studies

Item	Mean(d)	SD(d)	CDIF	NCDIF	Sig.
1	-0.02700	0.01300	0.00176	0.00090	ns
2	-0.14700	0.07500	0.00216	0.02715	0.001
3	0.07500	0.03600	-0.00274	0.00684	0.001
4	0.03900	0.02400	-0.00406	0.00207	0.05
5	-0.09000	0.05100	0.00882	0.01065	0.001
6	-0.06800	0.03200	0.00383	0.00557	0.01
7	-0.00800	0.00400	0.00028	0.00007	ns
8	0.06700	0.04200	0.00153	0.00626	0.001
9	0.07400	0.04000	-0.00013	0.00715	0.001
10	-0.00800	0.00500	-0.00019	0.00009	ns
11	0.07600	0.05100	0.00280	0.00846	0.001
12	0.02200	0.01200	-0.00014	0.00064	ns
13	-0.03900	0.01900	0.00306	0.00186	0.05
			Differential test functioning (DTF):		0.01697

APPENDIX H

P-VALUES DIFFERENCES FOR FIELD TEST AND OPERATIONAL ITEMS

Grade 3 ELA

Field Test Form	Field Test Position	Operational Position	Field Test P-Value	Operational P-Value	Difference
A	10	1	0.8990	0.8782	0.0208
C	54	2	0.6851	0.8252	-0.1402
A	8	3	0.7542	0.7488	0.0054
A	1	4	0.5792	0.5541	0.0251
C	51	5	0.8604	0.9207	-0.0603
A	11	6	0.6959	0.6889	0.0070
A	60	7	0.6393	0.7194	-0.0801
B	28	8	0.7313	0.7972	-0.0659
A	14	9	0.7002	0.7114	-0.0112
C	16	10	0.6324	0.5833	0.0491
C	15	11	0.6957	0.6006	0.0951
A	17	12	0.6340	0.6603	-0.0263
A	2	13	0.8634	0.8356	0.0278
B	10	14	0.8272	0.8377	-0.0105
A	20	15	0.7250	0.7287	-0.0037
C	21	16	0.6385	0.6277	0.0107
B	24	17	0.5454	0.5084	0.0370
A	23	18	0.7288	0.7196	0.0093
A	3	19	0.8306	0.7734	0.0572
A	5	20	0.6773	0.6139	0.0634
C	24	21	0.7755	0.7794	-0.0039
A	32	22	0.5176	0.5415	-0.0239
A	40	23	0.6515	0.6279	0.0237
A	24	24	0.7561	0.7604	-0.0043
B	55	25	0.4514	0.5008	-0.0494
C	3	26	0.6501	0.5605	0.0896
A	36	27	0.8145	0.8466	-0.0322
A	21	28	0.7016	0.7055	-0.0039
B	31	29	0.5538	0.5512	0.0026
B	57	30	0.6217	0.7241	-0.1024
C	29	31	0.5065	0.3618	0.1447
A	38	32	0.6436	0.6310	0.0126
A	26	33	0.3246	0.2985	0.0261
C	39	34	0.7591	0.7603	-0.0013
A	41	35	0.7956	0.8217	-0.0260

A	4	36	0.3702	0.3395	0.0307
A	44	37	0.6830	0.7481	-0.0650
A	37	38	0.4853	0.4923	-0.0070
A	54	39	0.6795	0.7163	-0.0368
A	47	40	0.6600	0.6926	-0.0325
A	48	41	0.6341	0.6986	-0.0645
A	50	42	0.5419	0.5506	-0.0088
B	58	43	0.6061	0.6320	-0.0259
B	45	44	0.7039	0.7441	-0.0402
A	53	45	0.5484	0.5549	-0.0065
A	35	46	0.7172	0.7022	0.0150
A	43	47	0.4971	0.4180	0.0791
B	60	48	0.5338	0.5604	-0.0266
A	16	49	0.5112	0.5039	0.0073
A	28	50	0.6596	0.7491	-0.0895
Mean:			**0.6539**	**0.6581**	

Grade 3 Math

Field Test Form	Field Test Position	Operational Position	Field Test P-Value	Operational P-Value	Difference
A	5	1	0.7682	0.7518	0.0164
A	49	2	0.8484	0.9379	-0.0895
B	7	3	0.8413	0.8298	0.0115
A	9	4	0.6487	0.6848	-0.0361
C	11	5	0.5598	0.5943	-0.0346
A	55	6	0.6738	0.7257	-0.0519
A	6	7	0.8669	0.8605	0.0064
A	13	8	0.8349	0.7932	0.0417
C	20	9	0.9457	0.9492	-0.0035
B	54	10	0.5610	0.5798	-0.0188
A	70	11	0.8418	0.9126	-0.0709
A	17	12	0.6629	0.6979	-0.0351
A	47	13	0.6701	0.6763	-0.0062
B	50	14	0.8928	0.9336	-0.0408
B	18	15	0.8336	0.8201	0.0135
A	21	16	0.6781	0.5870	0.0910
C	59	17	0.8297	0.8456	-0.0159
B	46	18	0.8151	0.9104	-0.0953
C	26	19	0.6165	0.6170	-0.0005
A	25	20	0.4739	0.4350	0.0390
C	68	21	0.7457	0.5582	0.1875
B	10	22	0.8363	0.6925	0.1438
A	3	23	0.9490	0.9002	0.0488
A	29	24	0.8075	0.8490	-0.0415
C	24	25	0.5643	0.5343	0.0300
C	67	26	0.6365	0.7153	-0.0788
B	61	27	0.6390	0.6348	0.0042
B	65	28	0.5781	0.5633	0.0147
B	26	29	0.7406	0.7050	0.0356
C	39	30	0.4321	0.4174	0.0147
B	66	31	0.6096	0.5088	0.1008
A	15	32	0.3734	0.3411	0.0323
A	1	33	0.8064	0.7886	0.0179
A	40	34	0.7913	0.8022	-0.0109
C	36	35	0.8742	0.8470	0.0272
A	62	36	0.5122	0.5263	-0.0141
C	62	37	0.6950	0.7438	-0.0487
A	44	38	0.7307	0.7616	-0.0309
B	31	39	0.6087	0.5297	0.0791
A	30	40	0.6111	0.5853	0.0258
A	48	41	0.7307	0.7957	-0.0650
B	59	42	0.8681	0.8941	-0.0259
A	67	43	0.8573	0.9069	-0.0496
B	51	44	0.8439	0.8529	-0.0090

A	52	45	0.6827	0.6894	-0.0067
A	16	46	0.7654	0.6868	0.0786
A	39	47	0.6650	0.6459	0.0191
A	56	48	0.6238	0.6377	-0.0139
B	4	49	0.9235	0.8519	0.0716
B	37	50	0.7414	0.7814	-0.0400
C	55	51	0.7918	0.7854	0.0064
A	60	52	0.7276	0.7369	-0.0093
A	65	53	0.3155	0.3131	0.0024
A	34	54	0.7912	0.7982	-0.0070
A	64	55	0.6481	0.6600	-0.0119
C	37	56	0.7492	0.6771	0.0721
B	53	57	0.7495	0.6014	0.1480
A	51	58	0.4673	0.4618	0.0055
C	16	59	0.7704	0.7141	0.0563
A	28	60	0.7684	0.7391	0.0293
Mean:			**0.7148**	**0.7063**	

Grade 3 Social Studies

Field Test Form	Field Test Position	Operational Position	Field Test P-Value	Operational P-Value	Difference
A	34	1	0.5051	0.5107	-0.0056
C	69	2	0.4872	0.6117	-0.1245
A	3	3	0.6358	0.5605	0.0753
A	5	4	0.5786	0.5346	0.0440
A	53	5	0.7882	0.8406	-0.0524
C	55	6	0.7025	0.7629	-0.0604
B	49	7	0.6510	0.7194	-0.0685
A	9	8	0.6196	0.5520	0.0677
B	61	9	0.6549	0.7703	-0.1154
C	59	10	0.5617	0.6558	-0.0941
A	13	11	0.6748	0.5834	0.0915
B	20	12	0.5834	0.6036	-0.0202
A	67	13	0.7079	0.6988	0.0091
C	19	14	0.7204	0.7744	-0.0540
A	17	15	0.4555	0.4589	-0.0034
A	12	16	0.7322	0.6756	0.0566
C	35	17	0.4446	0.4783	-0.0336
A	21	18	0.7104	0.6858	0.0246
C	33	19	0.5578	0.5880	-0.0303
A	69	20	0.7526	0.7802	-0.0277
A	26	21	0.6185	0.5779	0.0406
A	25	22	0.7674	0.7431	0.0243
A	28	23	0.4821	0.4502	0.0319
B	55	24	0.3898	0.4086	-0.0189
A	29	25	0.6178	0.6064	0.0114
B	68	26	0.7372	0.8708	-0.1336
A	14	27	0.7111	0.6535	0.0576
A	19	28	0.6931	0.5521	0.1410
C	70	29	0.3768	0.4196	-0.0427
A	68	30	0.5108	0.4421	0.0687
A	35	31	0.6078	0.5832	0.0246
C	68	32	0.5088	0.4946	0.0143
A	16	33	0.8169	0.7926	0.0243
A	40	34	0.6861	0.6601	0.0261
A	1	35	0.4758	0.4259	0.0500
B	24	36	0.3960	0.4164	-0.0204
C	30	37	0.6219	0.6357	-0.0137
A	44	38	0.6098	0.5896	0.0202
A	23	39	0.7173	0.6812	0.0361
C	24	40	0.6202	0.6642	-0.0440
A	48	41	0.4849	0.4975	-0.0126
C	62	42	0.4017	0.4681	-0.0664
B	66	43	0.4403	0.4381	0.0021
B	14	44	0.4335	0.4913	-0.0579

A	52	45	0.7090	0.6725	0.0365
C	2	46	0.8819	0.8308	0.0511
C	27	47	0.6243	0.5987	0.0256
A	56	48	0.7115	0.6556	0.0559
B	39	49	0.6842	0.6965	-0.0124
A	27	50	0.7926	0.7235	0.0691
B	28	51	0.5557	0.5390	0.0168
A	60	52	0.6643	0.6838	-0.0195
B	63	53	0.7020	0.7487	-0.0467
B	59	54	0.4103	0.5246	-0.1143
A	64	55	0.6667	0.6553	0.0114
A	70	56	0.7110	0.6896	0.0214
A	51	57	0.8855	0.8365	0.0489
B	19	58	0.6632	0.6628	0.0004
A	38	59	0.7603	0.7817	-0.0214
A	4	60	0.8806	0.7388	0.1418
Mean:			**0.6259**	**0.6241**	

Grade 5 ELA

Field Test Form	Field Test Position	Operational Position	Field Test P-Value	Operational P-Value	Difference
B	3	1	0.5744	0.6002	-0.0258
B	40	2	0.5847	0.5489	0.0359
C	52	3	0.6826	0.7948	-0.1122
C	7	4	0.8372	0.8521	-0.0149
C	58	5	0.8020	0.9539	-0.1518
C	2	6	0.8678	0.8889	-0.0211
B	8	7	0.7691	0.7909	-0.0218
B	57	8	0.5437	0.6630	-0.1193
C	49	9	0.4976	0.5067	-0.0090
A	11	10	0.6881	0.7089	-0.0208
B	49	11	0.6145	0.6120	0.0026
A	14	12	0.8380	0.8488	-0.0108
C	12	13	0.8649	0.8568	0.0080
A	10	14	0.4828	0.5012	-0.0184
A	17	15	0.4507	0.4590	-0.0083
A	22	16	0.7721	0.7913	-0.0192
A	20	17	0.7989	0.8386	-0.0398
B	22	18	0.6087	0.5826	0.0261
A	45	19	0.7251	0.7745	-0.0495
A	23	20	0.6798	0.7215	-0.0417
C	30	21	0.4299	0.4988	-0.0688
A	13	22	0.5841	0.6168	-0.0327
B	58	23	0.5376	0.5032	0.0344
C	18	24	0.3885	0.3827	0.0058
A	32	25	0.5410	0.5268	0.0142
A	59	26	0.4233	0.4222	0.0011
B	1	27	0.7811	0.7736	0.0075
A	34	28	0.7359	0.7755	-0.0396
C	22	29	0.5653	0.5964	-0.0311
C	29	30	0.6692	0.7021	-0.0328
C	13	31	0.5685	0.5670	0.0015
C	38	32	0.5085	0.5136	-0.0051
C	19	33	0.7995	0.7976	0.0019
C	27	34	0.4560	0.4692	-0.0132
A	41	35	0.5533	0.5373	0.0160
A	6	36	0.4262	0.4376	-0.0114
A	44	37	0.6878	0.7379	-0.0501
A	36	38	0.6428	0.6667	-0.0239
A	33	39	0.5700	0.5188	0.0511
A	47	40	0.6373	0.6158	0.0215
A	37	41	0.6429	0.6498	-0.0070
A	50	42	0.5740	0.6497	-0.0758
B	5	43	0.6760	0.5479	0.1281
C	31	44	0.6735	0.6487	0.0248

A	53	45	0.4949	0.5446	-0.0497
C	56	46	0.3612	0.3975	-0.0363
C	37	47	0.7334	0.7129	0.0206
A	3	48	0.8062	0.7482	0.0580
A	2	49	0.7851	0.7725	0.0126
A	60	50	0.7514	0.7504	0.0011
Mean:			**0.6337**	**0.6475**	

Grade 5 Math

Field Test Form	Field Test Position	Operational Position	Field Test P-Value	Operational P-Value	Difference
C	61	1	0.4756	0.6418	-0.1661
B	1	2	0.6213	0.6934	-0.0721
C	62	3	0.5867	0.7318	-0.1451
A	5	4	0.7457	0.8032	-0.0576
A	54	5	0.3752	0.2547	0.1205
B	32	6	0.6254	0.7102	-0.0849
B	38	7	0.7609	0.8102	-0.0493
A	9	8	0.6818	0.6928	-0.0110
C	32	9	0.4468	0.6000	-0.1532
B	7	10	0.6053	0.6418	-0.0365
A	13	11	0.6903	0.7165	-0.0263
B	58	12	0.3933	0.4994	-0.1061
C	57	13	0.4194	0.6525	-0.2330
A	67	14	0.6006	0.6081	-0.0075
A	17	15	0.5377	0.6191	-0.0813
A	37	16	0.3315	0.3673	-0.0358
C	47	17	0.4113	0.5054	-0.0941
A	21	18	0.5573	0.5701	-0.0129
A	53	19	0.3414	0.4042	-0.0628
B	59	20	0.6287	0.7448	-0.1160
A	31	21	0.6121	0.7622	-0.1500
A	25	22	0.4652	0.5221	-0.0570
A	24	23	0.5544	0.6449	-0.0905
B	46	24	0.7465	0.7585	-0.0120
A	29	25	0.6575	0.7172	-0.0597
C	67	26	0.5150	0.5556	-0.0407
A	39	27	0.5065	0.5093	-0.0027
B	15	28	0.9037	0.8905	0.0133
C	12	29	0.5205	0.4907	0.0298
B	43	30	0.5248	0.4623	0.0625
B	61	31	0.5879	0.7092	-0.1213
B	55	32	0.6950	0.7630	-0.0680
C	59	33	0.8008	0.9068	-0.1060
A	40	34	0.5089	0.6443	-0.1354
A	47	35	0.7267	0.7797	-0.0530
A	35	36	0.3602	0.4512	-0.0910
C	54	37	0.6110	0.7223	-0.1113
A	44	38	0.4718	0.6148	-0.1430
C	30	39	0.3658	0.4200	-0.0542
B	39	40	0.7003	0.7018	-0.0015
A	48	41	0.7701	0.8037	-0.0336
B	10	42	0.5900	0.6115	-0.0215
C	14	43	0.6947	0.6745	0.0202
A	6	44	0.8501	0.8301	0.0199

A	52	45	0.6831	0.7563	-0.0732
A	61	46	0.7599	0.8494	-0.0895
B	20	47	0.7539	0.7646	-0.0107
A	56	48	0.5243	0.5341	-0.0098
A	7	49	0.6011	0.6490	-0.0479
C	3	50	0.7142	0.6896	0.0246
C	34	51	0.4922	0.4837	0.0084
A	60	52	0.5226	0.5967	-0.0741
A	57	53	0.7726	0.7842	-0.0116
C	49	54	0.7216	0.8234	-0.1018
A	64	55	0.4802	0.5324	-0.0522
A	8	56	0.6887	0.6859	0.0029
B	70	57	0.4119	0.4736	-0.0617
C	50	58	0.3885	0.4390	-0.0505
C	63	59	0.4419	0.5213	-0.0794
C	2	60	0.7575	0.7364	0.0210
Mean:			**0.5882**	**0.6422**	

Grade 5 Social Studies

Field Test Form	Field Test Position	Operational Position	Field Test P-Value	Operational P-Value	Difference
A	16	1	0.2102	0.8390	-0.6288
B	4	2	0.7003	0.7416	-0.0413
B	53	3	0.5732	0.7005	-0.1273
A	5	4	0.3992	0.4739	-0.0748
C	58	5	0.5316	0.7295	-0.1979
A	15	6	0.5904	0.6389	-0.0485
A	70	7	0.5895	0.7289	-0.1394
A	9	8	0.5776	0.6309	-0.0533
B	27	9	0.6955	0.8239	-0.1284
A	41	10	0.5749	0.6185	-0.0436
A	13	11	0.4282	0.4666	-0.0384
A	8	12	0.2253	0.2778	-0.0525
B	7	13	0.5829	0.5185	0.0644
B	1	14	0.5045	0.5424	-0.0379
A	17	15	0.3232	0.4443	-0.1211
C	51	16	0.3705	0.4837	-0.1131
A	11	17	0.9325	0.9377	-0.0053
A	21	18	0.4770	0.6237	-0.1467
B	15	19	0.5045	0.6148	-0.1103
A	61	20	0.3556	0.4394	-0.0838
B	69	21	0.4598	0.5617	-0.1019
A	25	22	0.7167	0.7733	-0.0566
B	51	23	0.3268	0.4375	-0.1107
B	61	24	0.2221	0.3060	-0.0839
A	29	25	0.4159	0.4789	-0.0630
A	55	26	0.3239	0.3409	-0.0170
B	24	27	0.5142	0.6188	-0.1046
A	31	28	0.2908	0.3499	-0.0590
B	67	29	0.4291	0.5082	-0.0791
C	59	30	0.4852	0.6426	-0.1574
A	45	31	0.2601	0.2557	0.0044
A	7	32	0.4026	0.4531	-0.0505
C	55	33	0.2762	0.3045	-0.0283
A	40	34	0.5642	0.5677	-0.0035
B	28	35	0.2675	0.2986	-0.0311
B	57	36	0.4437	0.5339	-0.0903
B	68	37	0.2570	0.2982	-0.0412
A	44	38	0.6142	0.6562	-0.0419
B	30	39	0.4723	0.5900	-0.1177
A	42	40	0.3469	0.3393	0.0076
A	48	41	0.4827	0.5152	-0.0325
A	68	42	0.3178	0.3496	-0.0318
C	18	43	0.4855	0.4953	-0.0097
B	55	44	0.3426	0.3773	-0.0346

A	52	45	0.7036	0.7362	-0.0326
C	45	46	0.4392	0.5117	-0.0725
A	54	47	0.6016	0.6489	-0.0474
A	56	48	0.5041	0.6017	-0.0976
B	42	49	0.3666	0.4098	-0.0432
C	4	50	0.8646	0.8072	0.0574
C	26	51	0.4346	0.5730	-0.1384
A	60	52	0.5185	0.5832	-0.0647
C	61	53	0.4478	0.5140	-0.0661
A	19	54	0.4929	0.4837	0.0092
A	64	55	0.7154	0.7888	-0.0734
A	51	56	0.5355	0.5826	-0.0471
A	30	57	0.6167	0.6331	-0.0164
A	24	58	0.6315	0.6497	-0.0182
B	2	59	0.5952	0.4522	0.1429
A	4	60	0.9128	0.8744	0.0384
Mean:			**0.4874**	**0.5530**	

Grade 7 ELA

Field Test Form	Field Test Position	Operational Position	Field Test P-Value	Operational P-Value	Difference
C	24	1	0.6559	0.7315	-0.0757
B	55	2	0.8403	0.9345	-0.0941
C	3	3	0.7538	0.7919	-0.0381
A	21	4	0.7165	0.7582	-0.0417
C	58	5	0.8213	0.9315	-0.1102
C	59	6	0.7176	0.9107	-0.1930
A	8	7	0.7429	0.7748	-0.0319
A	16	8	0.5316	0.4757	0.0559
A	9	9	0.5244	0.5779	-0.0536
A	11	10	0.4690	0.5416	-0.0726
B	29	11	0.7590	0.8427	-0.0837
A	14	12	0.3837	0.4928	-0.1091
B	36	13	0.6265	0.6663	-0.0398
A	37	14	0.6810	0.7309	-0.0499
A	17	15	0.7788	0.8274	-0.0487
B	6	16	0.7494	0.7702	-0.0208
A	20	17	0.4584	0.5094	-0.0509
B	18	18	0.7467	0.7996	-0.0529
C	52	19	0.7675	0.8330	-0.0655
A	23	20	0.4878	0.4837	0.0041
A	33	21	0.8255	0.8315	-0.0060
C	12	22	0.6243	0.6775	-0.0532
B	39	23	0.7416	0.8223	-0.0807
C	15	24	0.4020	0.4822	-0.0803
A	15	25	0.6872	0.6828	0.0044
C	25	26	0.6374	0.6930	-0.0556
B	16	27	0.5512	0.6343	-0.0831
B	1	28	0.7232	0.7790	-0.0558
B	46	29	0.5296	0.5728	-0.0432
C	31	30	0.2563	0.3831	-0.1269
B	26	31	0.7846	0.8035	-0.0189
A	38	32	0.5305	0.5801	-0.0497
C	34	33	0.7342	0.7960	-0.0617
B	28	34	0.6350	0.7203	-0.0853
A	41	35	0.6832	0.7625	-0.0793
A	24	36	0.8147	0.7986	0.0161
A	44	37	0.5703	0.6658	-0.0954
A	32	38	0.5811	0.5680	0.0130
A	60	39	0.4017	0.4559	-0.0542
A	47	40	0.7047	0.7556	-0.0508
B	42	41	0.6110	0.6541	-0.0431
A	50	42	0.6907	0.7647	-0.0739
C	37	43	0.5500	0.6149	-0.0649
C	39	44	0.6325	0.6999	-0.0675

A	53	45	0.5172	0.5728	-0.0556
A	45	46	0.5202	0.5426	-0.0225
C	21	47	0.5586	0.6419	-0.0832
B	52	48	0.4909	0.5636	-0.0727
C	35	49	0.4357	0.4592	-0.0235
A	6	50	0.3420	0.5851	-0.2432
Mean:			**0.6196**	**0.6790**	

Grade 7 Math

Field Test Form	Field Test Position	Operational Position	Field Test P-Value	Operational P-Value	Difference
B	37	1	0.3374	0.4468	-0.1094
B	70	2	0.5939	0.6609	-0.0670
C	27	3	0.6762	0.7076	-0.0315
A	5	4	0.4296	0.4739	-0.0443
A	43	5	0.5860	0.6522	-0.0662
C	61	6	0.6236	0.7292	-0.1056
B	30	7	0.4130	0.4420	-0.0290
A	9	8	0.6516	0.6519	-0.0002
B	34	9	0.3288	0.2918	0.0370
A	3	10	0.6102	0.6009	0.0093
A	13	11	0.5965	0.6239	-0.0274
A	45	12	0.7370	0.8136	-0.0766
C	3	13	0.4456	0.4286	0.0171
B	28	14	0.6017	0.6151	-0.0134
A	17	15	0.6267	0.6561	-0.0294
A	24	16	0.5854	0.7536	-0.1682
C	24	17	0.8374	0.8885	-0.0510
A	21	18	0.5269	0.5770	-0.0501
C	33	19	0.3555	0.3997	-0.0442
C	34	20	0.5716	0.6626	-0.0909
B	15	21	0.7550	0.7276	0.0274
A	25	22	0.6529	0.7220	-0.0691
B	6	23	0.4284	0.4054	0.0230
B	4	24	0.5883	0.6126	-0.0243
A	29	25	0.5532	0.5997	-0.0464
A	18	26	0.6487	0.7154	-0.0666
B	62	27	0.4508	0.4683	-0.0174
A	50	28	0.7149	0.7220	-0.0071
C	58	29	0.6193	0.6324	-0.0131
A	70	30	0.5463	0.6105	-0.0642
B	53	31	0.4835	0.3828	0.1007
B	3	32	0.8085	0.7894	0.0191
A	7	33	0.5037	0.5230	-0.0193
A	40	34	0.7266	0.7696	-0.0430
C	63	35	0.7279	0.7208	0.0071
C	8	36	0.6377	0.6395	-0.0018
A	62	37	0.3747	0.4035	-0.0288
A	44	38	0.5389	0.5640	-0.0250
A	67	39	0.3632	0.5349	-0.1717
C	51	40	0.5242	0.5462	-0.0220
A	48	41	0.6217	0.6318	-0.0101
A	32	42	0.6833	0.8030	-0.1197
B	31	43	0.5666	0.6283	-0.0616
B	12	44	0.1616	0.2311	-0.0695

A	52	45	0.4930	0.4825	0.0105
A	36	46	0.4136	0.5417	-0.1280
A	4	47	0.4904	0.5338	-0.0434
A	56	48	0.5743	0.5887	-0.0145
B	46	49	0.3701	0.3780	-0.0079
B	10	50	0.7540	0.6740	0.0800
C	6	51	0.3620	0.4112	-0.0492
A	60	52	0.4262	0.4358	-0.0096
C	57	53	0.4084	0.3771	0.0313
A	27	54	0.6788	0.6530	0.0258
A	64	55	0.4085	0.4737	-0.0653
A	28	56	0.6343	0.6550	-0.0207
C	50	57	0.3216	0.5745	-0.2529
C	45	58	0.2535	0.1943	0.0593
B	69	59	0.2748	0.2195	0.0553
B	1	60	0.2849	0.2944	-0.0095
Mean:			**0.5327**	**0.5658**	

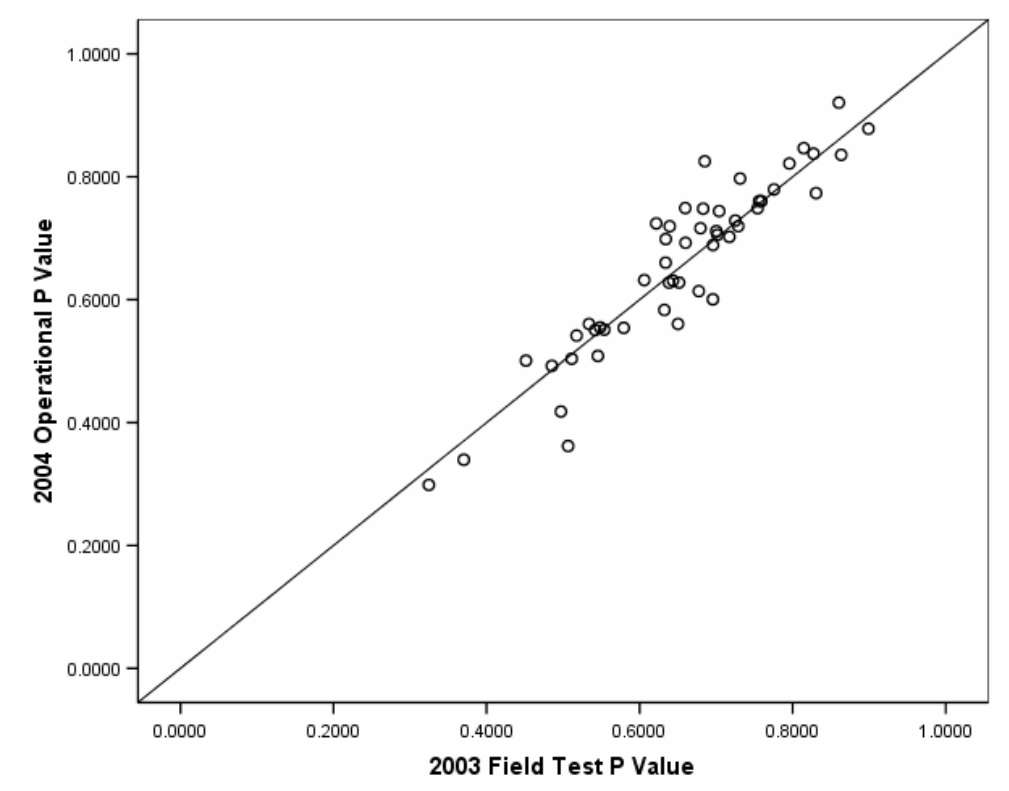
Field Test Form	Field Test Position	Operational Position	Field Test P-Value	Operational P-Value	Difference
A	11	1	0.4503	0.5484	-0.0981
B	12	2	0.4450	0.4715	-0.0264
A	41	3	0.4954	0.6061	-0.1107
A	5	4	0.4708	0.4777	-0.0069
B	58	5	0.3998	0.4121	-0.0123
B	39	6	0.4973	0.4716	0.0257
A	10	7	0.3444	0.3379	0.0065
A	9	8	0.6289	0.5482	0.0807
A	42	9	0.5420	0.4732	0.0688
B	35	10	0.5567	0.6364	-0.0797
A	13	11	0.4985	0.5729	-0.0744
C	43	12	0.3066	0.3594	-0.0528
A	27	13	0.4961	0.5547	-0.0586
A	26	14	0.7845	0.7912	-0.0066
A	17	15	0.3170	0.3250	-0.0079
B	19	16	0.5208	0.5261	-0.0053
A	66	17	0.3970	0.4350	-0.0380
A	21	18	0.3937	0.3489	0.0448
A	2	19	0.7599	0.6054	0.1545
B	66	20	0.5403	0.5570	-0.0167
A	51	21	0.4777	0.4138	0.0639
A	25	22	0.5053	0.4788	0.0266
A	7	23	0.4624	0.4300	0.0324
A	24	24	0.6511	0.5764	0.0747
A	29	25	0.5290	0.5415	-0.0125
A	57	26	0.5259	0.5282	-0.0023
B	11	27	0.3790	0.4208	-0.0418
A	23	28	0.4078	0.4111	-0.0033
A	50	29	0.8082	0.7881	0.0201
C	8	30	0.5791	0.6561	-0.0770
B	41	31	0.6524	0.6400	0.0124
B	46	32	0.6043	0.5855	0.0188
C	7	33	0.4714	0.4727	-0.0013
A	40	34	0.6253	0.6638	-0.0385
C	18	35	0.3724	0.4157	-0.0434
C	51	36	0.3675	0.5014	-0.1339
A	55	37	0.4862	0.4948	-0.0086
A	44	38	0.5715	0.6260	-0.0545
C	24	39	0.4288	0.5437	-0.1149
C	19	40	0.4835	0.6553	-0.1718
A	48	41	0.6555	0.6682	-0.0127
C	59	42	0.2657	0.3396	-0.0739
B	45	43	0.4697	0.4762	-0.0065
C	53	44	0.3676	0.4899	-0.1223

A	52	45	0.6510	0.7440	-0.0930
A	35	46	0.3867	0.3735	0.0132
C	35	47	0.3743	0.3901	-0.0158
A	56	48	0.5825	0.6190	-0.0365
C	39	49	0.3088	0.4188	-0.1101
C	36	50	0.6115	0.6998	-0.0883
B	62	51	0.5655	0.6176	-0.0521
A	60	52	0.4376	0.4146	0.0230
A	47	53	0.6808	0.6636	0.0172
B	8	54	0.6116	0.5009	0.1107
A	64	55	0.3975	0.4269	-0.0294
B	38	56	0.5566	0.4250	0.1316
B	37	57	0.7274	0.6918	0.0356
B	67	58	0.7147	0.7073	0.0074
A	28	59	0.4562	0.4606	-0.0044
B	42	60	0.3942	0.3774	0.0168
Mean:			**0.5075**	**0.5235**	

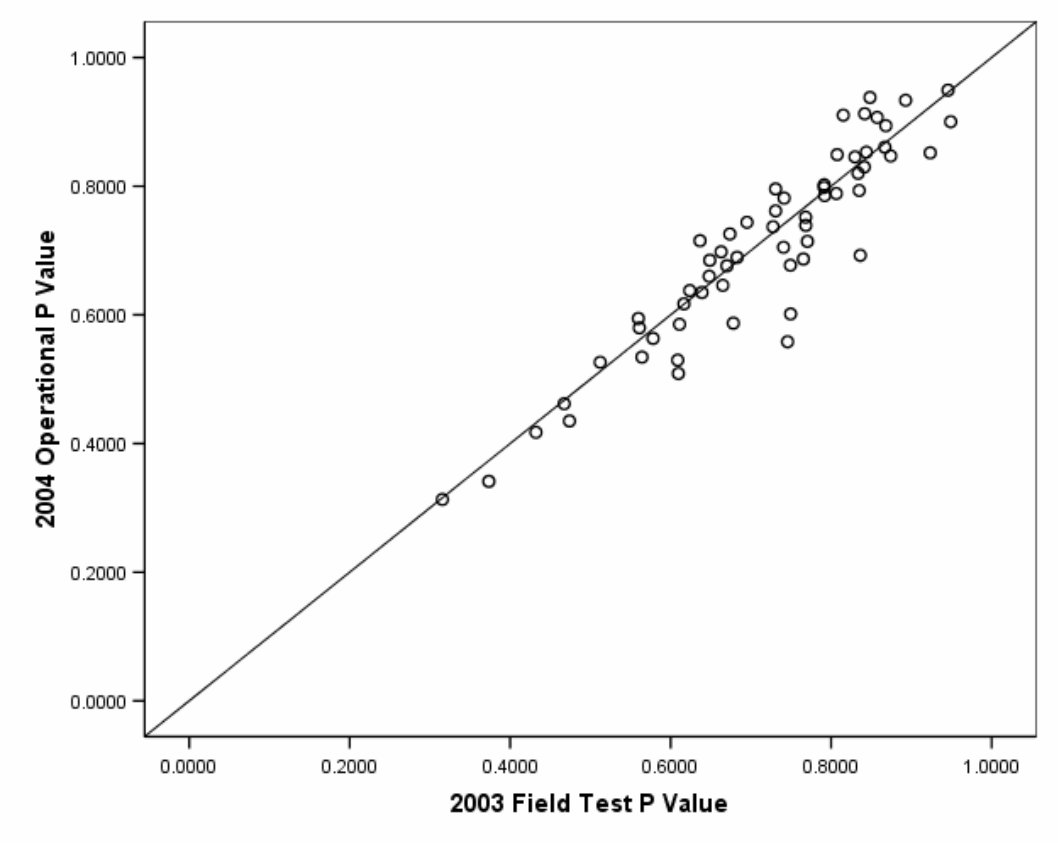
APPENDIX I

SCATTEPLOT OF P-VALUES DIFFERENCES FOR FIELD TEST AND OPERATIONAL ITEMS

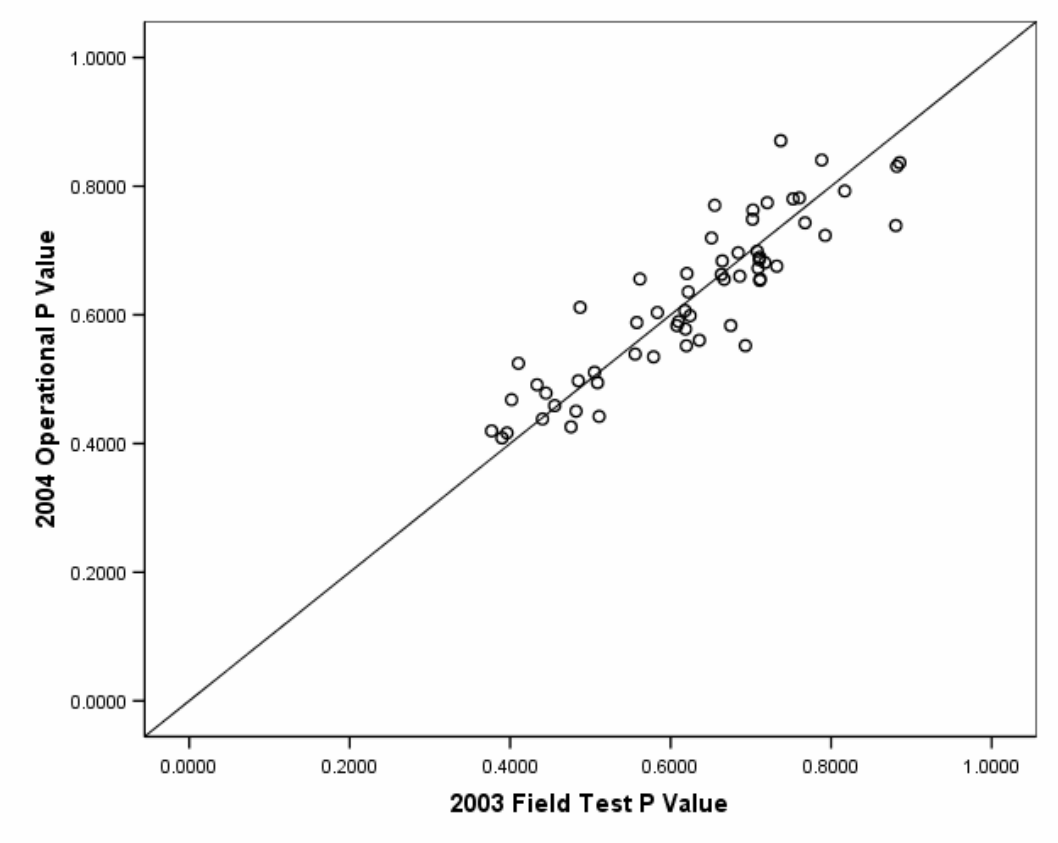
Grade 3 ELA



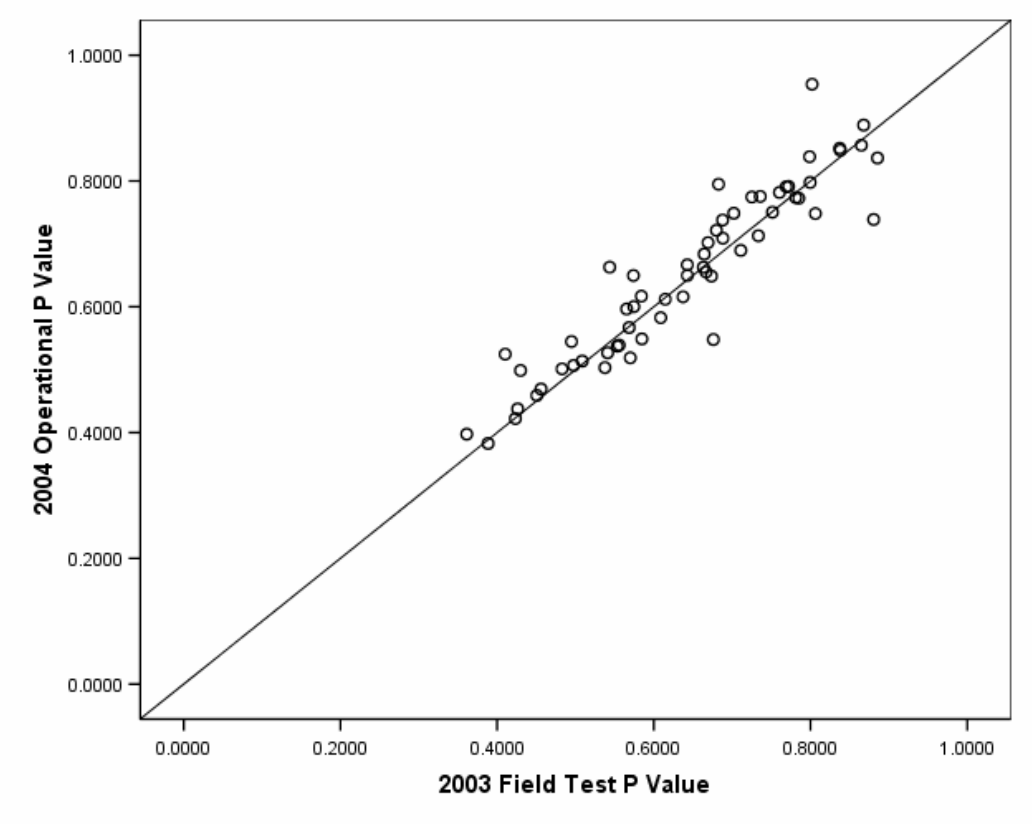
Grade 3 Math


2004 Operational P Value
1.0000
0.8000
0.6000
0.4000
0.2000
0.0000
0.0000
0.2000
0.4000
0.6000
0.8000
1.0000
2003 Field Test P Value

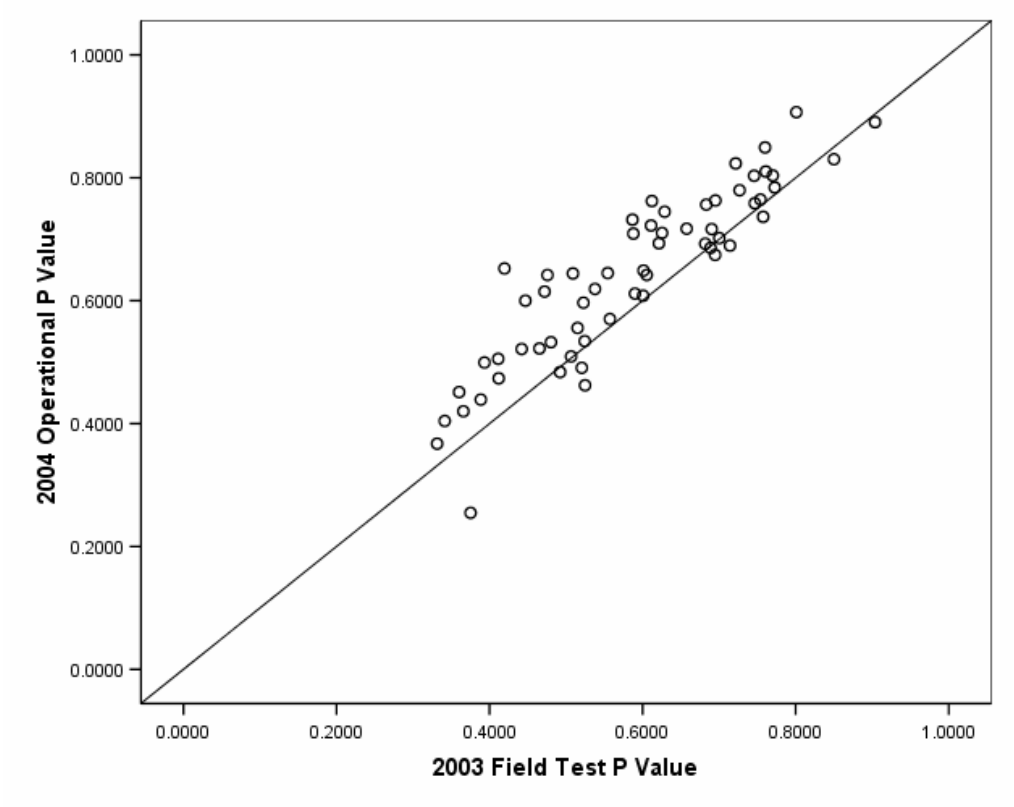
Grade 3 Social Studies



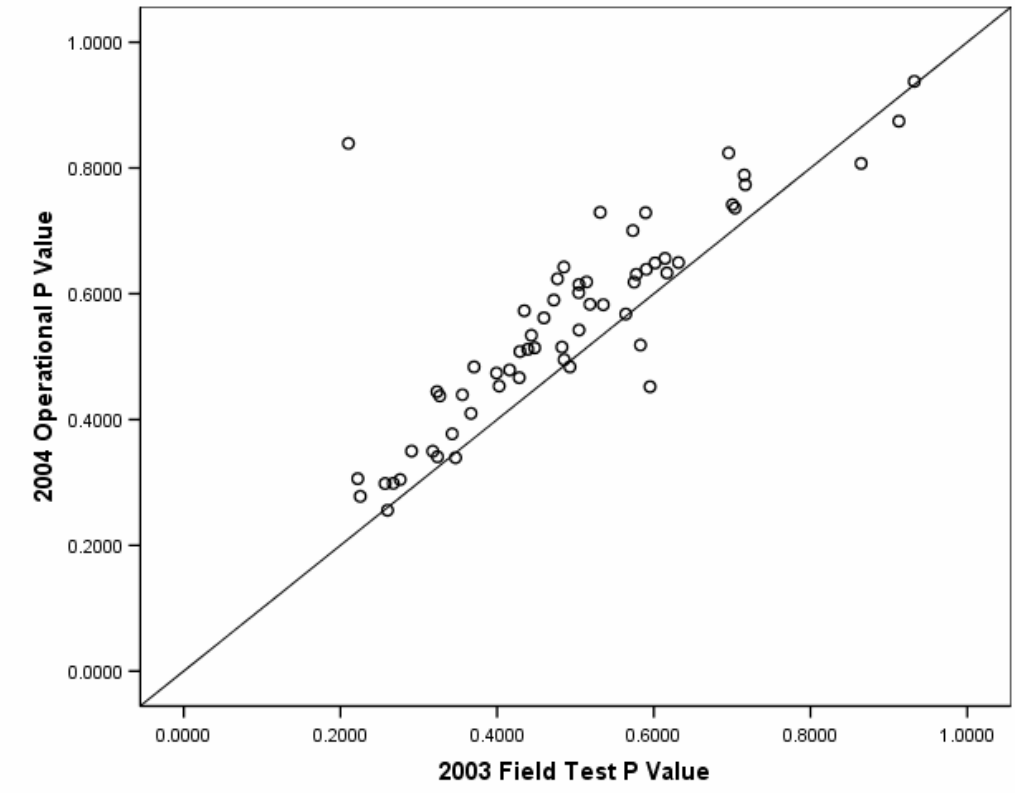
Grade 5 ELA


2004 Operational P Value
1.0000
0.8000
0.6000
0.4000
0.2000
0.0000
0.0000
0.2000
0.4000
0.6000
0.8000
1.0000
2003 Field Test P Value

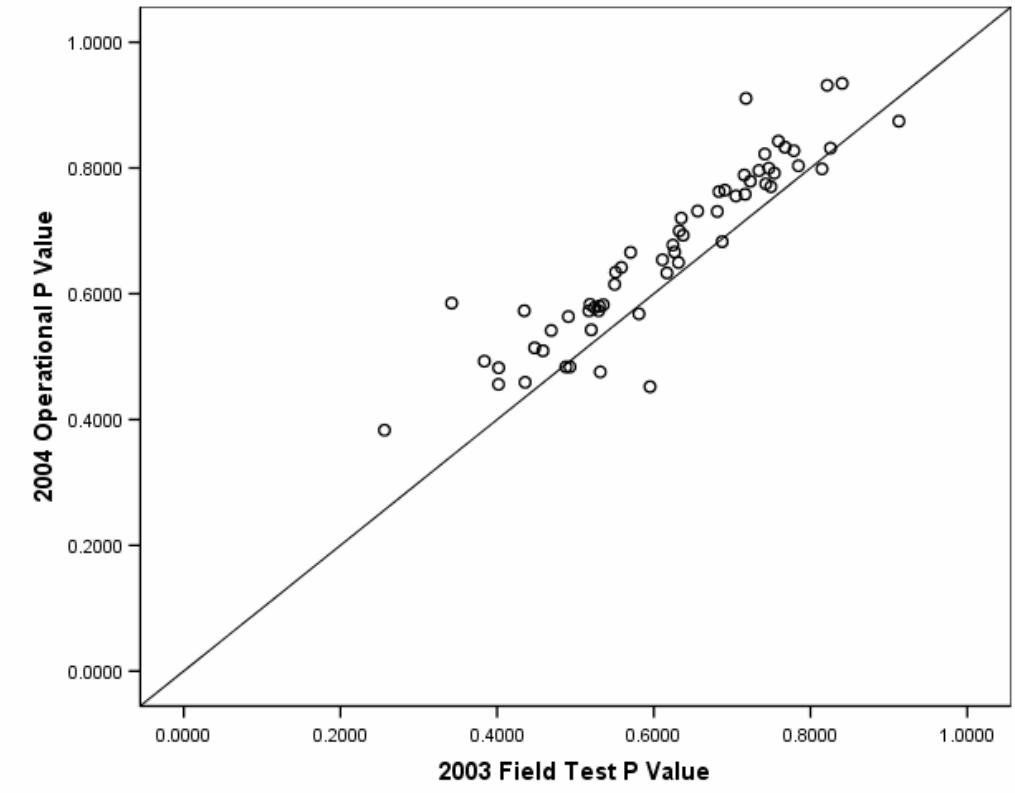
Grade 5 Math


2004 Operational P Value
1.0000
0.8000
0.6000
0.4000
0.2000
0.0000
0.0000
0.2000
0.4000
0.6000
0.8000
1.0000
2003 Field Test P Value

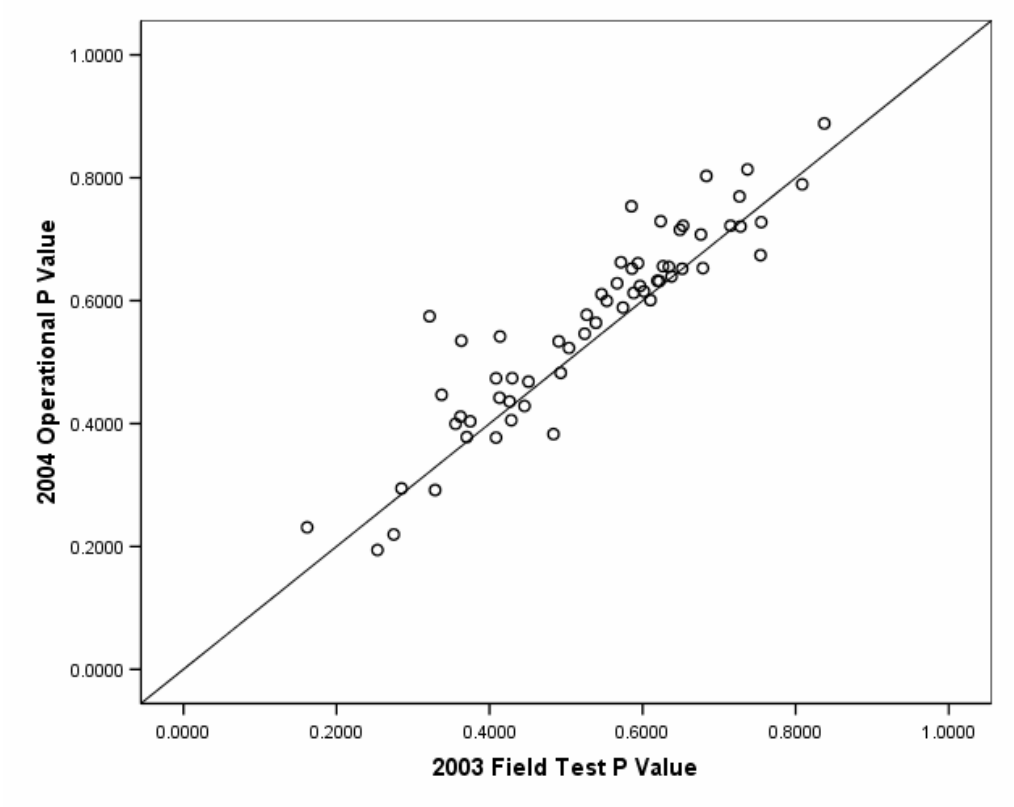
Grade 5 Social Studies



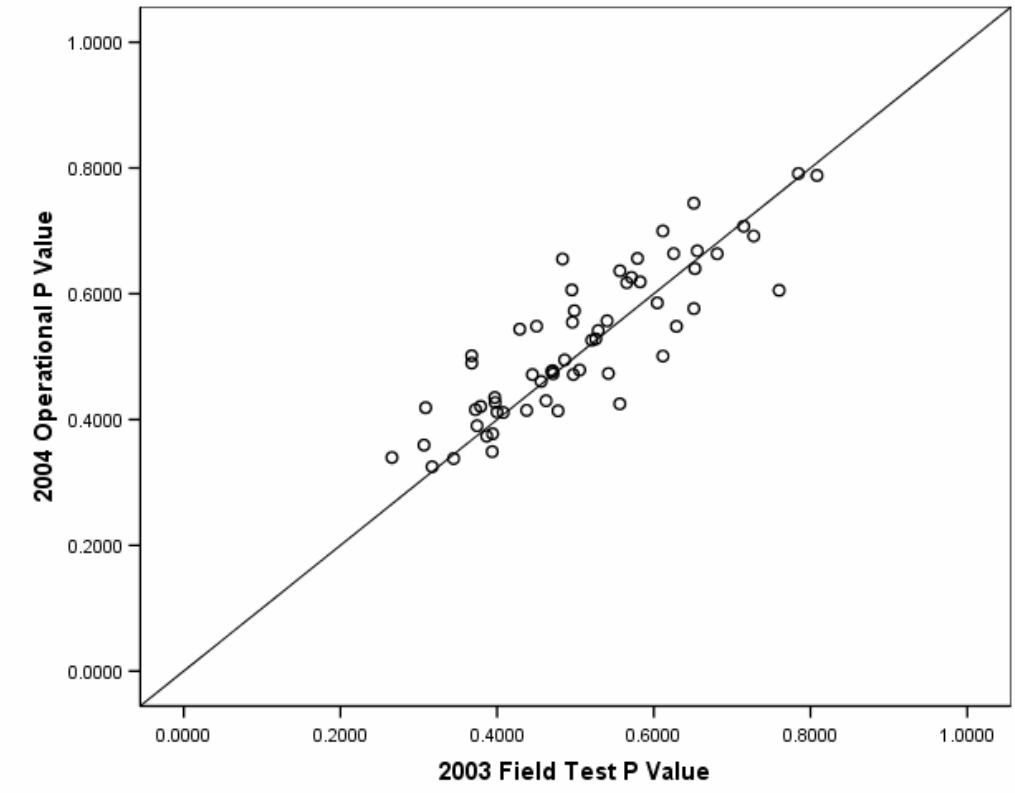
Grade 7 ELA


2004 Operational P Value
1.0000
0.8000
0.6000
0.4000
0.2000
0.0000
0.0000
0.2000
0.4000
0.6000
0.8000
1.0000
2003 Field Test P Value

Grade 7 Math


2004 Operational P Value
1.0000
0.8000
0.6000
0.4000
0.2000
0.0000
0.0000
0.2000
0.4000
0.6000
0.8000
1.0000
2003 Field Test P Value

Grade 7 Social Studies



APPENDIX J

ITEM PARAMETER DIFFERENCES FOR THE ONE, TWO, AND THREE PARAMETER LOGISTIC MODEL FOR GRADE THREE SOCIAL STUDIES

Grade 3, Social Studies 1PL			
Item	Pre-Equated Difficulty	Post-Equated Difficulty	Difference
1	0.31068	-0.04357	0.35424
2	0.32122	-0.54355	0.86477
3	-0.36838	-0.28376	-0.08462
4*	-0.18645	-0.15559	-0.03086
5	-1.26388	-1.96068	0.69680
6	-1.02766	-1.39183	0.36417
7	-0.79072	-1.12735	0.33662
8*	-0.56835	-0.24172	-0.32663
9	-0.79559	-1.44025	0.64465
10	-0.12092	-0.76788	0.64696
11*	-0.71453	-0.39805	-0.31648
12	-0.39536	-0.49871	0.10334
13	-0.71103	-1.00995	0.29892
14	-1.14027	-1.46957	0.32930
15*	0.15637	0.21948	-0.06311
16	-0.91465	-0.87718	-0.03747
17	0.57746	0.12263	0.45483
18*	-0.58611	-0.93094	0.34483
19	-0.09229	-0.41652	0.32423
20	-1.00836	-1.50053	0.49217
21	-0.26050	-0.36635	0.10585
22*	-1.11018	-1.25827	0.14809
23	0.42529	0.26026	0.16502
24	0.71254	0.46497	0.24756
25*	-0.54445	-0.51309	-0.03136
26	-1.33138	-2.22686	0.89547
27	-0.80238	-0.75102	-0.05136
28	-0.68682	-0.23912	-0.44770
29	0.99692	0.41233	0.58459
30	0.29578	0.29806	-0.00229
31	-0.21036	-0.39759	0.18723
32	0.20160	0.04190	0.15970
33	-1.51205	-1.58787	0.07582
34*	-0.72324	-0.78844	0.06520
35	0.44927	0.37839	0.07087

36	0.67005	0.42537	0.24467
37	-0.48117	-0.66759	0.18642
38*	-0.63788	-0.42887	-0.20901
39	-0.82382	-0.90515	0.08133
40	-0.47447	-0.81308	0.33861
41*	-0.25770	0.02778	-0.28548
42	0.83842	0.17038	0.66804
43	0.42882	0.31942	0.10939
44	0.45179	0.05807	0.39371
45*	-0.88936	-0.85260	-0.03676
46	-2.65620	-1.86630	-0.78990
47	-0.49773	-0.46917	-0.02856
48*	-0.64022	-0.76574	0.12552
49	-1.01406	-0.99494	-0.01913
50	-1.29774	-1.14816	-0.14958
51	-0.24457	-0.17239	-0.07219
52*	-0.40600	-0.91764	0.51164
53	-1.08572	-1.29333	0.20760
54	0.59976	-0.10172	0.70147
55*	-0.65453	-0.75873	0.10420
56	-0.75644	-0.94460	0.18816
57	-2.09371	-1.91130	-0.18241
58	-0.88837	-0.79878	-0.08960
59	-1.08163	-1.50832	0.42669
60	-2.08093	-1.23294	-0.84799
		Root Mean Square Difference=**.3763**	

* = Link Item

Grade 3, Social Studies 2PL						
Item	Pre-Equated Slope	Post-Equated Slope	Slope Difference	Pre-Equated Difficulty	Post-Equated Difficulty	Difficulty Difference
1	0.8510	0.7830	0.0680	0.2580	-0.0730	0.3310
2	0.4940	0.5170	-0.0230	0.1060	-0.6180	0.7240
3	0.3000	0.4480	-0.1480	-0.8490	-0.3740	-0.4750
4*	0.5098	0.4590	0.0508	-0.2150	-0.2160	0.0010
5	0.9000	0.8220	0.0780	-0.9470	-1.5600	0.6130
6	0.7680	0.4400	0.3280	-0.8050	-1.7330	0.9280
7	0.8710	0.7060	0.1650	-0.6260	-1.0070	0.3810
8*	0.5003	0.4570	0.0433	-0.6479	-0.3190	-0.3289
9	0.8910	0.5750	0.3160	-0.6230	-1.4630	0.8400
10	0.5740	0.5350	0.0390	-0.2610	-0.8380	0.5770
11*	0.8982	0.8560	0.0422	-0.5786	-0.3480	-0.2306
12	0.6290	0.5090	0.1200	-0.4260	-0.5760	0.1500
13	0.7830	0.6380	0.1450	-0.5880	-0.9670	0.3790
14	1.1090	1.0670	0.0420	-0.7190	-1.0450	0.3260
15*	0.5583	0.5410	0.0173	0.1542	0.1950	-0.0408
16	0.5750	0.5090	0.0660	-0.9550	-0.9870	0.0320
17	0.3360	0.3090	0.0270	0.4650	0.1700	0.2950
18*	0.8262	0.8820	-0.0558	-0.4964	-0.7450	0.2486
19	0.7810	0.7470	0.0340	-0.1950	-0.3860	0.1910
20	0.7800	0.8240	-0.0440	-0.8310	-1.2090	0.3780
21	0.5760	0.6290	-0.0530	-0.3020	-0.3780	0.0760
22*	0.9321	0.9360	-0.0039	-0.8687	-0.9590	0.0903
23	0.2490	0.3310	-0.0820	0.5110	0.3730	0.1380
24	0.5630	0.4600	0.1030	0.5100	0.5170	-0.0070
25*	0.3427	0.4580	-0.1153	-0.8347	-0.6390	-0.1957
26	1.3580	0.9970	0.3610	-0.8110	-1.5840	0.7730
27	0.7910	0.7900	0.0010	-0.6580	-0.6440	-0.0140
28	0.7710	0.6930	0.0780	-0.5750	-0.2460	-0.3290
29	0.4330	0.3940	0.0390	0.7990	0.5210	0.2780
30	0.3490	0.3190	0.0300	0.2620	0.4460	-0.1840
31	0.7030	0.7480	-0.0450	-0.1980	-0.3700	0.1720
32	0.7150	0.4970	0.2180	-0.0030	0.0140	-0.0170
33	0.8100	1.0980	-0.2880	-1.2020	-1.1090	-0.0930
34*	0.5430	0.4590	0.0840	-0.7763	-0.9620	0.1857
35	0.5730	0.6530	-0.0800	0.4140	0.3110	0.1030
36	0.3970	0.3930	0.0040	0.6530	0.5400	0.1130
37	1.1780	1.0520	0.1260	-0.3660	-0.5080	0.1420
38*	0.7950	0.7080	0.0870	-0.5485	-0.4080	-0.1405
39	0.8520	0.9600	-0.1080	-0.6430	-0.6980	0.0550
40	0.9240	0.7950	0.1290	-0.4030	-0.6910	0.2880
41*	0.2733	0.2470	0.0263	-0.4805	0.0230	-0.5035
42	0.3580	0.3020	0.0560	0.7510	0.2550	0.4960
43	0.6510	0.5560	0.0950	0.2250	0.2930	-0.0680
44	0.3680	0.4640	-0.0960	0.4340	0.0350	0.3990

45*	0.9123	0.9560	-0.0437	-0.7090	-0.6610	-0.0480
46	1.3530	1.4730	-0.1200	-1.3330	-1.1570	-0.1760
47	1.0790	0.9460	0.1330	-0.3880	-0.3860	-0.0020
48*	0.7510	0.8120	-0.0610	-0.5664	-0.6470	0.0806
49	1.2230	0.8650	0.3580	-0.6750	-0.8000	0.1250
50	0.9770	0.8410	0.1360	-0.9280	-0.9290	0.0010
51	0.1960	0.1960	0.0000	-0.6900	-0.4710	-0.2190
52*	0.7969	0.9780	-0.1811	-0.3552	-0.7010	0.3458
53	1.0600	0.8890	0.1710	-0.7470	-1.0100	0.2630
54	0.5030	0.6260	-0.1230	0.4580	-0.1310	0.5890
55*	1.0173	1.0720	-0.0547	-0.5068	-0.5670	0.0602
56	0.9970	1.0270	-0.0300	-0.5380	-0.7050	0.1670
57	0.9160	1.1500	-0.2340	-1.5220	-1.2900	-0.2320
58	0.4090	0.4930	-0.0840	-1.1020	-0.9230	-0.1790
59	0.6510	0.9260	-0.2750	-1.0140	-1.1430	0.1290
60	1.2910	1.2460	0.0450	-1.2730	-0.8390	-0.4340
	Root Mean Square Difference=.1367			Root Mean Square Difference =.3362		

* = Link Item

Grade 3 Social Studies
3PL

Item	Pre-Equated Slope	Post-Equated Slope	Slope Difference	Pre-Equated Difficulty	Post-Equated Difficulty	Difficulty Difference
1	1.90902	1.59056	0.31846	0.6284	0.41819	0.21021
2	0.80279	0.67949	0.12329	0.65628	0.01795	0.63833
3	0.38847	0.62646	-0.23799	-0.03468	0.31268	-0.34736
4*	0.73087	0.70957	0.0213	0.33349	0.44731	-0.11382
5	1.04017	0.84721	0.19296	-0.5918	-1.22855	0.63675
6	0.97611	0.49165	0.48446	-0.3124	-1.09669	0.7843
7	1.21363	0.83074	0.38289	-0.11827	-0.5101	0.39182
8*	0.63683	0.66983	-0.033	-0.08053	0.35715	-0.43768
9	1.11297	0.62972	0.48325	-0.16785	-0.95373	0.78587
10	0.82621	0.64935	0.17686	0.29009	-0.24517	0.53526
11*	1.28945	1.49509	-0.20564	-0.13579	0.1706	-0.30639
12	0.90126	0.67812	0.22314	0.14532	0.06338	0.08194
13	1.00122	0.76406	0.23716	-0.17175	-0.42819	0.25645
14	1.486	1.25137	0.23464	-0.27872	-0.63538	0.35666
15*	0.97732	1.10946	-0.13214	0.62626	0.70809	-0.08183
16	0.69218	0.61612	0.07606	-0.47149	-0.36133	-0.11016
17	0.56003	0.46997	0.09007	1.12183	0.98576	0.13607
18*	1.21133	1.17844	0.03289	-0.04102	-0.252	0.21098
19	1.36875	1.26288	0.10588	0.29869	0.16623	0.13247
20	0.95358	0.90676	0.04682	-0.41685	-0.79979	0.38294
21	0.80423	0.90327	-0.09904	0.20028	0.18278	0.0175
22*	1.11899	1.09296	0.02603	-0.49168	-0.53566	0.04398
23	0.40599	0.58051	-0.17452	1.25828	1.05799	0.20029
24	1.22223	1.03496	0.18727	0.92805	0.96134	-0.03329
25*	0.41667	0.58639	-0.16972	-0.10121	0.03714	-0.13835
26	1.63669	0.98422	0.65248	-0.43698	-1.32906	0.89208
27	1.05621	1.02302	0.03319	-0.20145	-0.14811	-0.05334
28	1.01622	1.13041	-0.11419	-0.12769	0.28963	-0.41732
29	1.13471	0.88226	0.25245	1.09635	1.00953	0.08681
30	0.52083	0.58972	-0.06889	0.87465	1.10982	-0.23517
31	1.10011	1.32784	-0.22773	0.27627	0.18929	0.08698
32	1.34328	0.91863	0.42465	0.47039	0.60245	-0.13206
33	0.92478	1.23448	-0.3097	-0.83094	-0.73667	-0.09427
34*	0.64888	0.55485	0.09403	-0.26778	-0.29846	0.03068
35	1.21451	1.69888	-0.48437	0.82491	0.72477	0.10014
36	0.89349	0.80227	0.09121	1.11289	1.05533	0.05756
37	1.82927	1.66063	0.16864	0.05484	-0.02807	0.08291
38*	1.08627	1.10444	-0.01817	-0.1014	0.14456	-0.24596
39	1.0799	1.22684	-0.14694	-0.24032	-0.24728	0.00696
40	1.3878	1.0368	0.35099	0.0612	-0.19243	0.25362
41*	0.35144	0.36262	-0.01118	0.40378	1.03389	-0.63011
42	0.65622	0.4576	0.19861	1.27436	1.07921	0.19515
43	1.9107	1.55338	0.35732	0.67441	0.74052	-0.0661
44	0.68098	0.77396	-0.09298	1.04243	0.64711	0.39531

45*	1.15736	1.23338	-0.07602	-0.31052	-0.2126	-0.09792
46	1.43588	1.59805	-0.16217	-1.04213	-0.85262	-0.18951
47	1.67688	1.54958	0.12729	0.05632	0.1024	-0.04608
48*	1.04738	1.15249	-0.10511	-0.09358	-0.11883	0.02525
49	1.6308	1.17279	0.45801	-0.25414	-0.28886	0.03473
50	1.13055	1.01693	0.11362	-0.59126	-0.46238	-0.12888
51	0.28067	0.26776	0.0129	0.61892	0.82649	-0.20757
52*	1.13518	1.31516	-0.17998	0.08223	-0.23235	0.31458
53	1.27654	1.01336	0.26318	-0.3321	-0.59159	0.25949
54	1.15802	1.13881	0.01921	0.90939	0.42267	0.48672
55*	1.61655	1.79936	-0.18281	-0.07166	-0.06714	-0.00452
56	1.37606	1.4363	-0.06024	-0.13003	-0.2316	0.10157
57	0.95959	1.21163	-0.25204	-1.27105	-0.98266	-0.2884
58	0.51963	0.59379	-0.07416	-0.3575	-0.29684	-0.06066
59	0.7569	1.04239	-0.28549	-0.57904	-0.73539	0.15635
60	1.29199	1.63979	-0.3478	-1.07151	-0.41272	-0.65879
	Root Mean Square Difference=.2360			**Root Mean Square Difference =.3250**		
Lower Asymptote: Pre-equated = .2025, Post Equated= .21393						

* = Link Item

APPENDIX K

PRE-EQUATED AND POST-EQUATED SCORE TABLES FOR ONE, TWO, AND THREE PARAMETER LOGISTIC MODEL FOR GRADE THREE SOCIAL STUDIES

Grade 3 Social Studies Score Table Comparison

1PL

	Pre-Equated		Post-Equated			
True Score	Theta	Scale Score	Theta	Scale Score	Theta Difference	SS Difference
0						
1	-3.3665	247	-3.4051	246	0.0386	1
2	-2.9082	258	-2.9700	256	0.0618	1
3	-2.6247	264	-2.7046	262	0.0799	2
4	-2.4147	269	-2.5088	267	0.0941	2
5	-2.2462	273	-2.3513	271	0.1051	2
6	-2.1044	276	-2.2180	274	0.1136	3
7	-1.9812	279	-2.1015	276	0.1203	3
8	-1.8718	282	-1.9974	279	0.1256	3
9	-1.7729	284	-1.9026	281	0.1297	3
10	-1.6824	286	-1.8152	283	0.1328	3
11	-1.5986	288	-1.7339	285	0.1353	3
12	-1.5203	290	-1.6574	287	0.1371	3
13	-1.4465	291	-1.5850	288	0.1385	3
14	-1.3765	293	-1.5160	290	0.1395	3
15	-1.3097	295	-1.4500	291	0.1403	3
16	-1.2457	296	-1.3865	293	0.1408	3
17	-1.1840	298	-1.3252	294	0.1412	3
18	-1.1243	299	-1.2658	296	0.1415	3
19	-1.0664	300	-1.2079	297	0.1415	3
20	-1.0099	302	-1.1516	298	0.1417	3
21	-0.9546	303	-1.0964	300	0.1418	3
22	-0.9004	304	-1.0423	301	0.1419	3
23	-0.8471	305	-0.9891	302	0.142	3
24	-0.7945	307	-0.9367	303	0.1422	3
25	-0.7426	308	-0.8849	304	0.1423	3
26	-0.6911	309	-0.8336	306	0.1425	3
27	-0.6399	310	-0.7828	307	0.1429	3
28	-0.5890	311	-0.7324	308	0.1434	3
29	-0.5382	312	-0.6821	309	0.1439	3
30	-0.4875	314	-0.6320	310	0.1445	3
31	-0.4367	315	-0.5819	311	0.1452	3
32	-0.3856	316	-0.5318	313	0.1462	3

33	-0.3343	317	-0.4816	314	0.1473	3
34	-0.2827	318	-0.4311	315	0.1484	3
35	-0.2305	320	-0.3802	316	0.1497	3
36	-0.1777	321	-0.3290	317	0.1513	4
37	-0.1242	322	-0.2772	319	0.153	4
38	-0.0699	323	-0.2247	320	0.1548	4
39	-0.0147	325	-0.1714	321	0.1567	4
40	0.0417	326	-0.1172	322	0.1589	4
41	0.0994	327	-0.0620	324	0.1614	4
42	0.1585	329	-0.0054	325	0.1639	4
43	0.2193	330	0.0526	326	0.1667	4
44	0.2819	331	0.1123	328	0.1696	4
45	0.3468	333	0.1740	329	0.1728	4
46	0.4141	335	0.2380	330	0.1761	4
47	0.4843	336	0.3047	332	0.1796	4
48	0.5579	338	0.3746	334	0.1833	4
49	0.6355	340	0.4484	335	0.1871	4
50	0.7179	342	0.5267	337	0.1912	4
51	0.8061	344	0.6107	339	0.1954	5
52	0.9015	346	0.7017	341	0.1998	5
53	1.0061	348	0.8018	344	0.2043	5
54	1.1227	351	0.9137	346	0.209	5
55	1.2558	354	1.0420	349	0.2138	5
56	1.4129	358	1.1940	353	0.2189	5
57	1.6078	362	1.3838	357	0.224	5
58	1.8720	368	1.6426	363	0.2294	5
59	2.3053	378	2.0703	373	0.235	5
60						

Root Mean Square Theta Difference: .1573

Grade 3 Social Studies Score Table Comparison

2PL

True Score	Pre-Equated Theta	Pre-Equated Scale Score	Post-Equated Theta	Post-Equated Scale Score	Theta Difference	SS Difference
0						
1	-5.0716	202	-5.4358	193	0.3642	9
2	-3.9515	229	-4.3449	219	0.3934	10
3	-3.3759	243	-3.7623	234	0.3864	9
4	-2.9998	252	-3.371	243	0.3712	9
5	-2.7239	259	-3.0787	250	0.3548	9
6	-2.507	264	-2.8459	256	0.3389	8
7	-2.3284	269	-2.6525	261	0.3241	8
8	-2.1765	272	-2.487	265	0.3105	8
9	-2.0439	276	-2.3419	268	0.298	7
10	-1.926	278	-2.2125	271	0.2865	7
11	-1.8194	281	-2.0953	274	0.2759	7
12	-1.7217	283	-1.9879	277	0.2662	6
13	-1.6311	286	-1.8883	279	0.2572	6
14	-1.5465	288	-1.7953	282	0.2488	6
15	-1.4667	290	-1.7077	284	0.241	6
16	-1.3909	291	-1.6246	286	0.2337	6
17	-1.3184	293	-1.5453	288	0.2269	6
18	-1.2487	295	-1.4691	290	0.2204	5
19	-1.1815	297	-1.3956	291	0.2141	5
20	-1.1161	298	-1.3243	293	0.2082	5
21	-1.0525	300	-1.2549	295	0.2024	5
22	-0.9902	301	-1.187	296	0.1968	5
23	-0.929	303	-1.1203	298	0.1913	5
24	-0.8687	304	-1.0546	300	0.1859	5
25	-0.8091	306	-0.9896	301	0.1805	4
26	-0.7499	307	-0.9251	303	0.1752	4
27	-0.6911	309	-0.8609	304	0.1698	4
28	-0.6323	310	-0.7967	306	0.1644	4
29	-0.5735	311	-0.7325	308	0.159	4
30	-0.5145	313	-0.668	309	0.1535	4
31	-0.4551	314	-0.6031	311	0.148	4
32	-0.3951	316	-0.5374	312	0.1423	3
33	-0.3344	317	-0.4709	314	0.1365	3
34	-0.2727	319	-0.4034	316	0.1307	3
35	-0.2098	320	-0.3345	317	0.1247	3
36	-0.1456	322	-0.2641	319	0.1185	3
37	-0.0798	323	-0.192	321	0.1122	3
38	-0.012	325	-0.1178	322	0.1058	3
39	0.0579	327	-0.0412	324	0.0991	2
40	0.1303	329	0.0382	326	0.0921	2
41	0.2057	330	0.1207	328	0.085	2
42	0.2843	332	0.2068	330	0.0775	2
43	0.3669	334	0.2971	333	0.0698	2

44	0.4539	336	0.3922	335	0.0617	2
45	0.546	339	0.4929	337	0.0531	1
46	0.6442	341	0.6001	340	0.0441	1
47	0.7496	344	0.715	343	0.0346	1
48	0.8633	346	0.8389	346	0.0244	1
49	0.9872	349	0.9736	349	0.0136	0
50	1.1235	353	1.1215	353	0.002	0
51	1.2749	356	1.2856	357	-0.0107	0
52	1.4456	361	1.4701	361	-0.0245	-1
53	1.6412	365	1.6808	366	-0.0396	-1
54	1.87	371	1.9266	372	-0.0566	-1
55	2.1453	378	2.2211	379	-0.0758	-2
56	2.4895	386	2.5877	388	-0.0982	-2
57	2.9457	397	3.0712	400	-0.1255	-3
58	3.6129	413	3.7748	417	-0.1619	-4
59	4.8192	443	5.0421	448	-0.2229	-5
60						

Root Mean Square Theta Difference: .2006

Grade 3 Social Studies Score Table Comparison

3PL

True Score	Pre-Equated Theta	Pre-Equated Scale Score	Post-Equated Theta	Post-Equated Scale Score	Theta Difference	SS Difference
0						
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						
13	-3.1126	242	-5.1135	196	2.0009	46
14	-2.3430	259	-2.9470	246	0.604	14
15	-1.9599	268	-2.3634	259	0.4035	9
16	-1.7045	274	-2.0212	267	0.3167	7
17	-1.5114	278	-1.7782	272	0.2668	6
18	-1.3547	282	-1.5886	277	0.2339	5
19	-1.2218	285	-1.4323	280	0.2105	5
20	-1.1055	288	-1.2984	283	0.1929	4
21	-1.0015	290	-1.1805	286	0.179	4
22	-0.9070	292	-1.0745	288	0.1675	4
23	-0.8199	294	-0.9778	290	0.1579	4
24	-0.7387	296	-0.8881	293	0.1494	3
25	-0.6624	298	-0.8043	294	0.1419	3
26	-0.5900	299	-0.7251	296	0.1351	3
27	-0.5210	301	-0.6498	298	0.1288	3
28	-0.4546	302	-0.5776	300	0.123	3
29	-0.3904	304	-0.5080	301	0.1176	3
30	-0.3280	305	-0.4405	303	0.1125	3
31	-0.2670	307	-0.3748	304	0.1078	2
32	-0.2072	308	-0.3105	306	0.1033	2
33	-0.1481	309	-0.2473	307	0.0992	2
34	-0.0895	311	-0.1849	309	0.0954	2
35	-0.0312	312	-0.1230	310	0.0918	2
36	0.0271	313	-0.0614	311	0.0885	2
37	0.0856	315	0.0003	313	0.0853	2
38	0.1446	316	0.0623	314	0.0823	2
39	0.2042	317	0.1247	316	0.0795	2
40	0.2648	319	0.1880	317	0.0768	2
41	0.3265	320	0.2524	318	0.0741	2
42	0.3897	322	0.3182	320	0.0715	2
43	0.4546	323	0.3858	321	0.0688	2

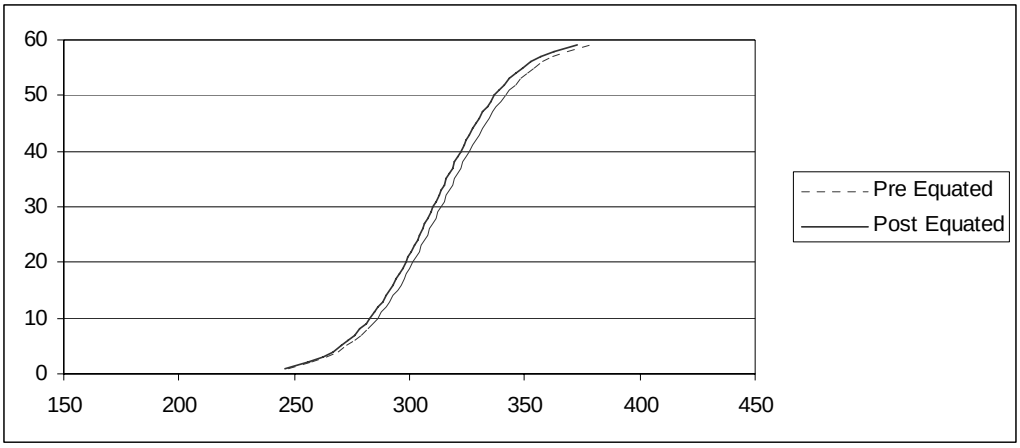
44	0.5216	325	0.4555	323	0.0661	2
45	0.5911	326	0.5278	325	0.0633	1
46	0.6636	328	0.6033	326	0.0603	1
47	0.7398	330	0.6825	328	0.0573	1
48	0.8205	331	0.7664	330	0.0541	1
49	0.9068	333	0.8560	332	0.0508	1
50	1.0000	335	0.9527	334	0.0473	1
51	1.1022	338	1.0585	337	0.0437	1
52	1.2158	340	1.1760	339	0.0398	1
53	1.3445	343	1.3092	342	0.0353	1
54	1.4939	347	1.4637	346	0.0302	1
55	1.6725	351	1.6486	350	0.0239	1
56	1.8957	356	1.8791	355	0.0166	0
57	2.1934	363	2.1848	362	0.0086	0
58	2.6365	373	2.6341	373	0.0024	0
59	3.4679	392	3.4592	391	0.0087	0
60						

Root Mean Square Theta Difference: .3317

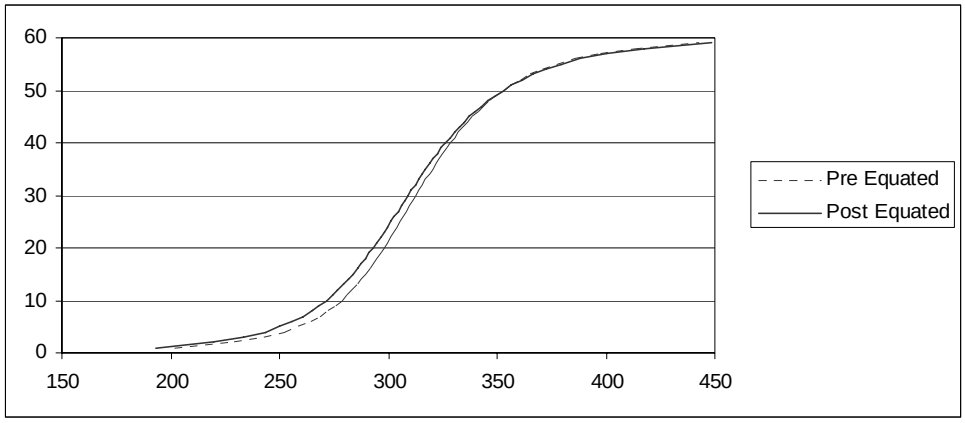
APPENDIX L

PRE-EQUATED AND POST-EQUATED SCORING GRAPHS FOR ONE, TWO, AND THREE PARAMETER LOGISTIC MODEL FOR GRADE THREE SOCIAL STUDIES

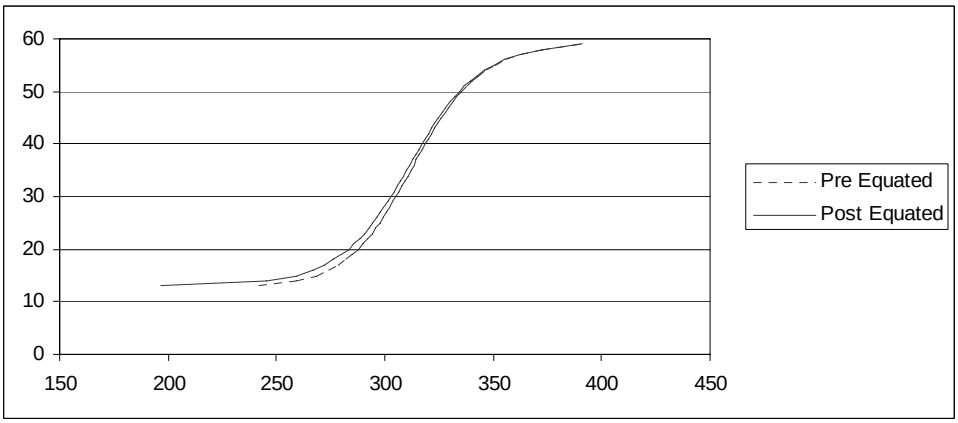
One Parameter Model



Two Parameter Model



Three Parameter Model



APPENDIX M

RESULTS OF DIFFERENTIAL ITEM AND TEST FUNCTIONING ANALYSES FOR ONE, TWO, AND THREE PARAMETER MODEL, GRADE THREE SOCIAL STUDIES

One Parameter Model

Item	Mean(d)	SD(d)	CDIF	NCDIF	Sig.
1	0.11600	0.04900	0.30460	0.01574	0.001
2	0.24700	0.09800	0.67040	0.07058	0.001
3	-0.00800	0.00400	-0.02224	0.00008	ns
4	0.02500	0.01100	0.06838	0.00073	ns
5	0.12800	0.12200	0.36020	0.03125	0.001
6	0.10100	0.07900	0.29086	0.01656	0.001
7	0.09400	0.06100	0.26993	0.01251	0.001
8	-0.12500	0.05500	-0.34857	0.01869	0.001
9	0.14100	0.09800	0.40406	0.02944	0.001
10	0.16200	0.07300	0.45348	0.03147	0.001
11	-0.09200	0.04500	-0.25990	0.01041	0.001
12	0.02200	0.01000	0.06171	0.00059	ns
13	0.06200	0.03600	0.17845	0.00519	0.001
14	0.06200	0.04900	0.17657	0.00616	0.001
15	-0.01800	0.00800	-0.04564	0.00037	ns
16	0.01800	0.01200	0.05233	0.00046	ns
17	0.07300	0.03500	0.18139	0.00659	0.001
18	0.10500	0.06000	0.30188	0.01468	0.001
19	0.11000	0.04800	0.30517	0.01437	0.001
20	0.09300	0.07000	0.26572	0.01350	0.001
21	0.05000	0.02200	0.13937	0.00299	0.01
22	0.02800	0.02000	0.07982	0.00118	ns
23	0.06700	0.03000	0.16996	0.00537	0.001
24	0.03300	0.01800	0.07920	0.00143	ns
25	-0.04000	0.01900	-0.11376	0.00199	0.05
26	0.13200	0.13300	0.37047	0.03522	0.001
27	-0.01200	0.00700	-0.03562	0.00020	ns
28	-0.10800	0.05100	-0.30366	0.01416	0.001
29	0.11900	0.06600	0.27628	0.01852	0.001
30	-0.00400	0.00200	-0.00961	0.00002	ns
31	0.07400	0.03300	0.20628	0.00655	0.001
32	0.05900	0.02500	0.15491	0.00407	0.01
33	0.02800	0.02600	0.07904	0.00146	ns
34	-0.01000	0.00600	-0.02928	0.00014	ns

35	0.00100	0.00100	0.00290	0.00000	ns
36	0.04500	0.02400	0.10959	0.00262	0.05
37	0.03700	0.01900	0.10618	0.00174	ns
38	-0.04300	0.02100	-0.12104	0.00226	0.05
39	0.02500	0.01500	0.07310	0.00088	ns
40	0.09400	0.04900	0.26819	0.01125	0.001
41	-0.07000	0.02900	-0.18997	0.00572	0.001
42	0.18000	0.08700	0.44096	0.03976	0.001
43	0.04300	0.02000	0.10907	0.00227	0.05
44	0.08600	0.03900	0.21956	0.00892	0.001
45	0.00300	0.00200	0.00818	0.00001	ns
46	-0.06200	0.07500	-0.16928	0.00945	0.001
47	-0.00700	0.00300	-0.02055	0.00006	ns
48	0.05000	0.02700	0.14190	0.00320	0.01
49	0.02400	0.01700	0.06987	0.00086	ns
50	-0.00600	0.00400	-0.01598	0.00005	ns
51	-0.04200	0.01800	-0.11470	0.00207	0.05
52	0.14300	0.07500	0.40789	0.02610	0.001
53	0.03800	0.02700	0.10786	0.00216	0.05
54	0.20000	0.08500	0.51411	0.04709	0.001
55	0.03700	0.02000	0.10648	0.00180	ns
56	0.04000	0.02400	0.11511	0.00216	0.05
57	-0.00300	0.00300	-0.00770	0.00002	ns
58	0.00800	0.00500	0.02315	0.00009	ns
59	0.06500	0.04900	0.18586	0.00663	0.001
60	-0.13400	0.12300	-0.38023	0.03303	0.001
			Differential test functioning (DTF):		6.7223

Two Parameter Model

Item	Mean(d)	SD(d)	CDIF	NCDIF	Sig.
1	0.08900	0.04400	0.18169	0.00990	0.001
2	0.13900	0.02700	0.23777	0.02009	0.001
3	-0.03300	0.03400	-0.09703	0.00225	ns
4	0.01700	0.01200	0.04280	0.00044	ns
5	0.13300	0.11200	0.35625	0.03032	0.001
6	0.10100	0.12500	0.32277	0.02573	0.001
7	0.07900	0.06500	0.21089	0.01046	0.001
8	-0.09700	0.02200	-0.16996	0.00988	0.001
9	0.11900	0.11900	0.34577	0.02836	0.001
10	0.09600	0.04000	0.19944	0.01082	0.001
11	-0.07300	0.03000	-0.09734	0.00622	0.001
12	0.00700	0.04100	0.06775	0.00176	ns
13	0.05300	0.05400	0.15553	0.00570	0.001
14	0.07900	0.05500	0.19399	0.00926	0.001
15	-0.01500	0.02100	0.00676	0.00065	ns
16	0.00900	0.05100	0.08299	0.00273	0.05
17	-0.00200	0.02000	0.02447	0.00039	ns
18	0.08100	0.04500	0.18326	0.00859	0.001
19	0.05700	0.02800	0.11998	0.00397	0.01
20	0.08300	0.05100	0.19722	0.00951	0.001
21	0.03200	0.02500	0.01623	0.00162	ns
22	0.02100	0.02400	0.06546	0.00104	ns
23	0.01400	0.03300	-0.02444	0.00130	ns
24	-0.01800	0.04100	0.02927	0.00202	ns
25	-0.03200	0.04100	-0.10606	0.00273	0.05
26	0.15700	0.16000	0.45448	0.05014	0.001
27	-0.00500	0.00700	0.00003	0.00008	ns
28	-0.08000	0.03200	-0.09553	0.00738	0.001
29	0.02300	0.00900	0.04574	0.00061	ns
30	-0.02800	0.02100	-0.01394	0.00123	ns
31	0.05500	0.02600	0.11601	0.00368	0.01
32	0.00200	0.08100	0.11543	0.00656	0.001
33	0.03300	0.03600	0.03356	0.00243	0.05
34	-0.01100	0.04900	0.05015	0.00257	0.05
35	-0.00700	0.03100	-0.05280	0.00102	ns
36	-0.00300	0.00600	0.00358	0.00004	ns
37	0.02600	0.02000	0.06416	0.00109	ns
38	-0.03500	0.02700	-0.02390	0.00196	ns
39	0.02700	0.01800	0.03139	0.00106	ns
40	0.06700	0.05000	0.16932	0.00691	0.001
41	-0.05400	0.01100	-0.06654	0.00300	0.05
42	0.07100	0.01300	0.12255	0.00520	0.01
43	-0.00600	0.05600	0.06812	0.00317	0.05
44	0.03400	0.03000	0.00990	0.00200	ns
45	0.00100	0.00600	-0.00609	0.00004	ns
46	-0.01900	0.01400	-0.04344	0.00056	ns

47	-0.00900	0.04200	0.04035	0.00189	ns
48	0.03700	0.01900	0.08071	0.00171	ns
49	0.03000	0.08100	0.14954	0.00742	0.001
50	0.00700	0.02900	0.04519	0.00087	ns
51	-0.03800	0.00500	-0.05127	0.00150	ns
52	0.11000	0.04100	0.18478	0.01386	0.001
53	0.04400	0.07200	0.16226	0.00706	0.001
54	0.11100	0.02800	0.13825	0.01315	0.001
55	0.02800	0.01300	0.05220	0.00094	ns
56	0.04200	0.02100	0.08518	0.00224	0.05
57	0.00700	0.01100	0.00669	0.00017	ns
58	0.01400	0.00800	0.03224	0.00026	ns
59	0.05600	0.04600	0.05147	0.00518	0.001
60	-0.10600	0.06500	-0.24149	0.01553	0.001
			Differential test functioning (DTF):		4.2637

Three Parameter Model

Item	Mean(d)	SD(d)	CDIF	NCDIF	Sig.
1	0.07700	0.06300	0.15781	0.00984	0.001
2	0.14100	0.04300	0.28396	0.02165	0.001
3	-0.04000	0.03300	-0.11859	0.00264	0.05
4	0.00900	0.00200	0.01867	0.00008	ns
5	0.13700	0.11600	0.40256	0.03204	0.001
6	0.11700	0.13400	0.39562	0.03169	0.001
7	0.10000	0.08700	0.30644	0.01741	0.001
8	-0.09600	0.03400	-0.19827	0.01041	0.001
9	0.13400	0.11900	0.41409	0.03206	0.001
10	0.10400	0.05100	0.25710	0.01350	0.001
11	-0.07000	0.04400	-0.11422	0.00680	0.001
12	0.02000	0.04500	0.10431	0.00241	0.05
13	0.04800	0.05300	0.16575	0.00517	0.001
14	0.09300	0.07700	0.27287	0.01463	0.001
15	-0.02000	0.02000	-0.00797	0.00082	ns
16	0.00900	0.05400	0.08917	0.00305	0.05
17	0.00500	0.02600	0.04751	0.00070	ns
18	0.08100	0.05700	0.22545	0.00991	0.001
19	0.05800	0.03600	0.13866	0.00471	0.01
20	0.08400	0.05700	0.23404	0.01031	0.001
21	0.02500	0.03200	-0.00358	0.00164	ns
22	0.02600	0.03100	0.09042	0.00163	ns
23	0.00800	0.02800	-0.02869	0.00087	ns
24	-0.01100	0.04000	0.03821	0.00172	ns
25	-0.03100	0.03500	-0.10673	0.00217	ns
26	0.17300	0.16800	0.54502	0.05799	0.001
27	0.00400	0.02300	0.03855	0.00053	ns
28	-0.07800	0.04100	-0.12150	0.00776	0.001
29	0.02500	0.03000	0.07259	0.00153	ns
30	-0.04000	0.01800	-0.04697	0.00193	ns
31	0.05200	0.02900	0.12509	0.00351	0.05
32	0.00100	0.07700	0.11894	0.00591	0.001
33	0.03200	0.03100	0.05303	0.00199	ns
34	-0.01200	0.04800	0.04420	0.00249	0.05
35	0.00500	0.03500	-0.02590	0.00123	ns
36	0.01500	0.02900	0.06347	0.00105	ns
37	0.03500	0.03200	0.10118	0.00227	0.05
38	-0.03700	0.02300	-0.05605	0.00193	ns
39	0.02900	0.01200	0.05796	0.00100	ns
40	0.07800	0.06700	0.23311	0.01059	0.001
41	-0.05700	0.01300	-0.08546	0.00342	0.05
42	0.07400	0.02600	0.15773	0.00622	0.001
43	-0.00300	0.05400	0.06621	0.00289	0.05
44	0.04800	0.02000	0.05554	0.00272	0.05
45	0.00500	0.00400	0.00480	0.00004	ns
46	-0.00400	0.01500	-0.00516	0.00025	ns

47	-0.00500	0.03500	0.03612	0.00123	ns
48	0.03800	0.02400	0.10133	0.00201	ns
49	0.03800	0.07500	0.17337	0.00708	0.001
50	0.00600	0.02700	0.04288	0.00079	ns
51	-0.02600	0.01600	-0.02413	0.00093	ns
52	0.10700	0.05100	0.22333	0.01411	0.001
53	0.05900	0.07400	0.21337	0.00902	0.001
54	0.12200	0.05600	0.19715	0.01802	0.001
55	0.02100	0.01500	0.02987	0.00064	ns
56	0.03800	0.02000	0.08381	0.00185	ns
57	0.00900	0.00700	0.01931	0.00013	ns
58	0.02900	0.02100	0.08054	0.00126	ns
59	0.05400	0.04100	0.07553	0.00455	0.01
60	-0.10100	0.06900	-0.27027	0.01488	0.001
			Differential test functioning (DTF):		5.4431