DOCUMENT RESUME

ED 445 060 TM 031 676

AUTHOR Ito, Kyoko; Sykes, Robert C.
TITLE An Evaluation of "Intentional" Weighting of Extended-Response or Constructed-Response Items in Tests with Mixed Item Types.
PUB DATE 2000-06-00
NOTE 95p.; Paper presented at the Annual National Conference of the Council of Chief State School Officers on Large-Scale Assessment (30th, Snowbird, UT, June 25-28, 2000).
PUB TYPE Numerical/Quantitative Data (110) -- Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS *Constructed Response; Elementary Education; Essay Tests; Test Construction; *Test Items; Test Length; Writing Tests
IDENTIFIERS *Weighting (Statistical)

ABSTRACT

This study investigated the practice of weighting a type of test item, such as constructed response, more than other types of items, such as selected response, to compute student scores for a mixed-item type of test. The study used data from statewide writing field tests in grades 3, 5, and 8 and considered two contexts, that in which a single extended response writing prompt is "intentionally" or "purposefully" weighted twice in computing student scores (ER context) and one in which a set of constructed response items, including one extended response item, is intentionally weighted twice (CR context). The weighting option was compared with the use of no weighting. In either context, the criterion for the two options (weighting and no weighting with a shorter form) was the administration of twice as many items as the items that are deliberately weighted, combined with no use of weighting. The three options were compared in terms of student scores as well as raw-score-to-scale score conversion tables. The state uses number-correct scoring as opposed to pattern scoring. Either intentionally weighted or un-weighted scores on the shorter form are, on average, very comparable to the criterion un-weighted scores from the longer form. On the level of individual student scores, as compared with the un-weighted shorter-form scores, more of the weighted shorter-form scores (2-5% more in the ER context and 1-9% more in the CR context) differ from the target scores by more than 10 points. The ramifications of the small decreases in individual score accuracy associated with purposeful weighting would depend on the purposes for which the scores are used and other factors. However, it is clear that purposeful weighting can never compensate for the loss of score accuracy caused by the shorter length of an actual test taken. Two appendixes contain a discussion of the item response models used in the study and a sample graph for one test item. (Contains 13 tables, 24 figures, and 12 references.) (Author/SLD)

An Evaluation of "Intentional" Weighting of Extended-Response or Constructed-Response Items in Tests With Mixed Item Types

Kyoko Ito, CTB/McGraw-Hill,
and
Robert C. Sykes, CTB/McGraw-Hill

A paper presented at the annual National Conference on Large Scale Assessment, Snowbird, Utah, June, 2000.

Abstract

The study investigated a fairly recent practice in some states that weights a type of item (e.g., constructed-response items) more than other types of items (e.g., selected-response items) to compute student scores for a mixed-item-type test. The study used data from statewide writing field-tests in three grades (3, 5, and 8) and considered two contexts: the context where a single extended-response writing prompt is "intentionally" or "purposefully" weighted twice in computing student scores ("ER context"), and the context where a set of constructed-response items, including one extended-response item, is intentionally weighted twice ("CR context"). The weighting option was compared against no use of such weighting. In either context, the criterion for the two options (weighting and no weighting with a shorter form) was the administration of twice as many items as the items that are deliberately weighted, combined with no use of weighting (no weighting with a longer form).

The three options were compared in terms of student scores as well as raw-score-to-scale-score conversion tables. The state uses number-correct scoring as opposed to pattern scoring. Either intentionally weighted or unweighted scores on the shorter form are, on average, very comparable to the criterion unweighted scores from the longer form. On the level of individual student scores, as compared with the unweighted shorter-form scores, more of the weighted shorter-form scores (2-5% more in the ER context, and 1-9% more in the CR context) differ from the target scores by more than ten points. The ramifications of the small decreases in individual score accuracy associated with purposeful weighting would depend on the purposes for which the scores are used and other factors. However, it is clear that purposeful weighting can never compensate for the loss of score accuracy caused by the shorter length of an actual test taken.

Introduction

Although weighting composite scores to obtain a weighted sum has been done for years (e.g., a weighted sum of verbal and quantitative scores), deliberately weighting one type of item in a test more than others is a recent practice. This practice has arisen in some assessment programs as more testing programs adopt tests containing mixed item types, such as selected-response (SR), short constructed-response (CR), and extended-response (ER) types. One of the motivations for the practice seems to be the desire to allow open-ended items to have a greater impact on the total test score. The present study investigated the impact of purposefully weighting a type of item in writing tests that contained all three item types. The kind of weighting investigated here is intentionally or purposefully imposed by human judgment, as opposed to other psychometrically-based weighting schemes (e.g., item pattern scoring based on item response theory, weighting based on reliability, weighting by the numbers of items).

The issue of intentional weighting was investigated in two situations: (a) a single extended writing prompt was weighted twice, so that the score points from extended writing were doubled ("ER context"), and (b) a set of short constructed-response and extended-response items were weighted twice ("CR context"). Take situation (a) as an example. A frequently used, "traditional" approach to attaining twice as many score points from extended writing as from a single prompt is to actually administer a second comparable prompt. Although this option of administering an additional item or items is more desirable from the perspective of the generalizability of total test scores, it may not be viable because of the increase in testing time. A possible alternative is to administer a single extended-response item and explicitly give it twice the weight, so the result would be the same as the conventional approach with regard to the sum of score points. Surely, one cannot expect every score, weighted or unweighted, from a shorter test with a single set of open-ended items to be very similar to the student's score from a longer test with twice as many open-ended items. However, given the wish to have open-ended items contribute more to the total test score, intentional weighting would be a viable alternative if it yields student scores that are fairly comparable to those obtained without such weighting when both are compared with scores from the conventional approach of "twice as many items."

The study compared the three options using live data from a state. The three options are:

- Option 1 (a longer form with no weighting) : requires administering twice as many ER and/or CR items, but involves no explicit weighting. This option honors the desire to weight an item or items of a selected type(s) by administering more items.
  [denoted in tables and figures "(unweighted) ER2" and "(unweighted) CR2," respectively, for the ER and CR contexts],
- Option 2 (a shorter form with double-weighting) : administers a single ER item or a single set of CR and ER items, and gives these items double weight in producing scores. This option also honors the wish to have an item or items weighted twice.
  [denoted in tables and figures "weighted ER1" and "weighted CR1," respectively, for the ER and CR contexts], and
- Option 3 (a shorter form with no weighting) : requires the administration of the same number of items as option 2, but involves no intentional weighting. Unlike option 1 or 2, this option disregards the desire to weight an item or items of a selected type(s).

[denoted "unweighted ER1" and "unweighted CR1," respectively, for the ER and CR contexts].

To reiterate, the target option is option 1, while the option of primary interest is option 2. The reason option 3 was added is that the option of interest (option 2) differs from the criterion option 1 in two aspects: test length and the use of purposeful weighting. Option 3, which differs from option 1 in only one aspect, test length, helps unravel the confounding. Any difference observed between option 1 and option 3 represents what is expected from the difference in form length. By comparing the option 2 – option 1 difference against the option 3 – option 1 difference, the effect of intentional weighting could be examined with the impact of test length removed.

The following summarizes all the option – context combinations:

Option	ER context	CR context
Option 1	(Unweighted) ER2	(Unweighted) CR2
Option 2	Weighted ER1	Weighted CR1
Option 3	Unweighted ER1	Unweighted CR1

In the study, the three options were compared in terms of raw-score-to-scale-score (RS-SS) conversion tables and the resulting scale scores. The testing program from which the tests and data came utilizes item response theory (IRT) to obtain the parameter estimates for the items and employs the number-correct scoring method in which each possible raw-score total is converted to a scale score based on a RS-SS table, which, in turn, is based on the item parameter estimates.

Method

Data Source

Data came from a field-test of Writing items for a state assessment program in grades 3, 5, and 8. The field tests contained two ER items and five, nine, or 15 CR items, depending on the grade, as well as 38 or 50 SR items. The field-test items were analyzed and calibrated using the IRT models described in Appendix A. From the pool of field-tested items, a set of items was selected for an operational form in each grade that met the content blueprint and the statistical criteria.

Test Forms Used in the Study

Regardless of the grade, the operational form contains a single ER item and was used with weighting as the "weighted ER1" form for option 2 and used with no weighting as the "unweighted ER1" form for option 3. The "(unweighted) ER2" form for option 1 was created by adding the other ER item, which resulted in two ER items. Attempts were made to make the two ER items as equally difficult as possible. The ER1 and ER2 forms have the identical set of SR and CR items.

Next, the "CR1" and "CR2" forms were constructed in such a way that the CR2 form has twice as many CR and ER items as the CR1 form and yet they are as comparable as possible in test difficulty. Because the CR2 form has more CR and ER items, it inevitably differs from the CR1

form in terms of internal consistency and content coverage. As in the ER context, the "(unweighted) CR2" form was used for option 1, whereas the CR1 form was used with weighting as the "weighted CR1" form for option 2 and used without weighting as the "unweighted CR1" form for option 3.

All the forms within a grade (i.e., ER1, ER2, CR1, and CR2) have the identical set of SR items. Tables 1 - 3 present lists of items included in each of the forms. The check mark (√) indicates that the item is in the form. The tables also show the number of students in the calibration sample, the mean p-values, and Feldt-Raju reliability values. The ER1 and ER2 forms, and the CR1 and CR2 forms are comparable in average test difficulty.

At the bottom of each table are the mean p-value for the ER and CR items that are in the CR2 form but not in the CR1 form, and the mean p-value for the ER and CR items that are in the CR1 form, that is, the ER and CR items that are doubly weighted. These mean p-values are within .02 of each other, suggesting the approximate equivalence in average difficulty between the additional items in the CR2 form and the doubly-weighted "virtual" items in the weighted CR1 form.

Table 4 presents further details of the forms. The table shows the numbers of SR, CR, and ER items, the total numbers of items and score points in the form, and the form's coverage of content by objective. Note that in the CR1 forms, the number of ER and CR items to be weighted twice increases over three grades. The grade 3 CR1 form has three ER/CR items to be weighted; the grade 5 CR1 form has 5 such items; the grade 8 CR1 form contains 7 such items.

Analyses

As noted before, the field-test items were calibrated to obtain the item parameter estimates. At each grade, the item parameter estimates from the field-test calibration were used as the true values in the comparisons of the forms. For example, after the CR2 form and the CR1 form were constructed from the field-test form, they were not re-calibrated. The same parameter estimates from the single field-test calibration were used for the same item in all study forms in which it appeared, whether the item was in the CR1 form or the CR2 form or the ER forms.[1] The parameter estimates were initially in the logit-like metric, but were placed onto a scale-score (SS) scale with a multiplier of 50 and an additive constant of 500. The highest and lowest scale scores of 200 and 800 were imposed on students' scale scores.

Weighting came into play when RS-SS tables were generated using these parameter estimates. All RS-SS tables were produced using CTB's proprietary software program named "FLUX". For option 1 with the ER2 or CR2 form that was always unweighted, a usual, unweighted RS-SS table was generated. For option 2 with weighting, a weighted RS-SS table was generated using the ER1 or CR1 form by giving double weight to a single ER item or to all the CR and ER items. For option 3 with no weighting, an unweighted RS-SS table was produced using the ER1 or CR1

[1] If the items had been re-calibrated for each study form, the re-calibrated parameter estimates for the items, say, in the CR2 form would likely be different from those for the items in the CR1 form. This is because the CR2 form contains more CR items than the CR1 form.

form without engaging weighting. Table 5 provides an overview of the forms in terms of which items were deliberately weighted and which items were intentionally left unweighted.

Using the RS-SS tables generated, students in the calibration samples were scored for each of the option – context combinations.

Results

ER context

The RS-SS tables for the ER context are provided in Tables 6 – 8. In Table 6, the RS-SS pairs for the grade 3 ER forms are listed for options 1 – 3, from left to right. Table 7 shows the comparable RS-SS pairs for the grade 5 ER forms, and Table 8 for the grade 8 ER forms.

Comparisons of the RS-SS tables and the test characteristic curves (TCCs)

The RS-SS pairs for the ER context are graphically depicted in Figures 1 –3 in terms of test characteristic curves (TCCs) with the percent correct, as opposed to the number correct score, on the Y axis. Although plots are often a great tool to evaluate differences, they usually do not show as much precision as the actual numbers. For example, in Figure 1, the option 1 and option 2 TCCs are so close, with the maximum SS difference of 3 points, that they are not differentiated in the figure. In those cases particularly, the RS-SS tables should be consulted.

The figures show that although the option 2 TCC is usually closer to the target option 1 TCC, both the option 2 and option 3 TCCs are so comparable to the criterion option 1 TCC that it seems reasonable to declare that both the weighted form (option 2) and the unweighted short form (option 3) are alternate forms of the unweighted long form with additional items (option 1). The close comparability of the TCCs among the three options is particularly notable in the SS range between 400 and 550. Despite the similarities, the option-3 TCCs are slightly but consistently higher than the option-1 TCCs on the low and high ends.

Comparisons in standard errors (SEs) and SE curves

In addition to the RS-SS pairs, the RS-SS tables (Tables 6 – 8) also show the standard errors (SEs) for the SSs. These are "constrained" standard errors. They are constrained in such a way that the upper or lower bound of the 1SE band in SS for a given RS is never above the upper or lower bound of the 1SE band in SS for (RS +1).

The SEs are plotted in Figures 4 – 6. The RS-SS tables and the SE plots show that

- in all cases, the SE curve for option 1 (ER2), as expected from longer form length, is the lowest throughout the SS range,
- as far as the low end of the SS range below 350 is concerned, the SE curve for option 2 (weighted ER1) is nearly identical or very similar to the criterion SE curve under option 1 (ER2). As compared with option 2, the SE curve for option 3 (unweighted ER1) is very slightly higher, that is, farther away from the target option-1 curve.
- for the mid-range of SSs between 350 and 600, the three SE curves for options 1 – 3 are either practically identical or very comparable.

- on the high end of the SS range above 600, although option 2 and option 3 are very similar in terms of SEs, the former has slightly higher SE than the latter. The SE plot for either option is substantially greater than that for the target SE for option 1. At a SS of 700, the linearly-interpolated SE for either option 2 or option 3 is about 10 to 20 SS points higher than that for the option 1 SE.

The weighted SE curve is lower on the low end of the range, indicating that the ER item being weighted twice is located in this very easy range. However, the p-value for the item, .58 in Table 1, suggests otherwise. The plot in Appendix B shows an unusual-looking TCC for the item, depicting a very low location. This phenomenon, which is not limited to this item, seems to be caused by the fact that in the item calibrations, the rounded average of six trait scores for the same prompt was used as the student's response to the item. All ER items in the state's tests were calibrated in this manner. Extended-response items are usually difficult. If the ER items had been scored in a typical fashion with no averaging of six trait scores, they would likely have lowered the SE curves on the high end of the SS range.

Scale score comparisions

The comparisons of options 2 and 3 against option 1 with regard to the RS-SS tables and curves have demonstrated both marked similarities and noticeable dissimilarities, particularly on the extreme ends of the SS range. The crucial question is: how do the similarities and dissimilarities translate into differences in students' scale scores? If the students are located in the SS range where the three options are very comparable, differences in the RS-SS tables may be inconsequential for most students.

To answer this question, for each grade, three sets of scale scores (SSs) for the students in the scaling sample were compared, that is, option 1 SSs, option 2 SSs, and option 3 SSs. The results of the scale-score comparisons in the ER context are summarized in Table 12.

Regardless of grade, the mean SS for either option 2 or option 3 is extremely similar to the criterion mean SS for option 1, indicating that the three options are tau-equivalent. The largest SS difference is half a scale-score point for the option 1 – option 3 comparison at grade 8. The SS standard deviations are very similar among the three options.

To compare the three options on the level of the individual student, two difference scores were computed for each student, one between the option 2 SS and the option 1 SS and the other between the option 3 SS and the option 1 SS. The means and standard deviations of the difference scores, presented in Table 12, display the same pattern described above. The table also shows the percentages of students with "relatively small" difference scores. Two types of "relatively small" difference scores are used: differences equal to or smaller than 5 SS points in absolute magnitude, and those equal to or smaller than 10 SS points in absolute magnitude. Since minimum standard errors are in a proximity of 15 and 20 SS points, the 5-point SS difference is roughly a quarter to a third of a minimum SE, while the 10-point SS difference is approximately a half to two-thirds of a minimum SE.

Across three grades, the weighted option-2 SSs are within 5 SS points of the target option-1 SSs in 60% - 66% of the students and within 10 SS points in 82% - 86% of the students. The

unweighted SSs are within 5 SS points of the criterion SSs in 61% – 80% of the students, and within 10 SS points in 86% - 89% of the students. Thus, for all three grades, either by the 5-point or 10-point criterion, the option-3 SSs are similar to the criterion option-1 SSs for more students than are the option-2 SSs, except when the 5-points criterion is used at grade 3.

The pairs of SSs are plotted in Figures 7 through 12 to see how close the SSs are among the three options. For example, Figure 7 plots the option 1 SSs against the option 2 SSs for grade 3 students, and Figure 8 plots the option 1 SSs against the option 3 SSs for the same students. Figure 9 is a plot of the option 1 SSs against the option 2 SSs for the grade 5 students, while Figure 10 is a plot of the option 1 SSs against the option 3 SSs for the same students.

In terms of correlations that are included in the plots, the option 3 SSs seem as similar to the target option 1 SSs as are the option 2 SSs. One may note while comparing Figures 7 and 8 that the scatterplot for the option 3 – option 1 SS pairs is slightly tighter along the approximate 45-degree line than for the option 2 – option 1 SS pairs, indicating that the option 3 SSs are slightly closer to the criterion option 1 SSs than the option 2 SSs are to the criterion SSs. This is also observed in the comparisons for the remaining two grades, that is, between Figures 9 vs. 10 and between Figures 11 vs.12. This characteristic is in line with the greater numbers of relatively accurate individual scores observed above for the no-weighting option (option 3), relative to the weighting option (option 2).

The plots also reveal that the option-1 and option-3 SSs are not tau-equivalent at the low (and high) ends of the SS range, indicating that the option-3 SSs are, on average, not very similar to the option-1 SSs at the extreme sections of the SS range. For example, the option-3 SSs for the students in a relatively low range tend to be lower, sometimes considerably lower, than the criterion SSs under option 1, but are rarely higher than the target SSs. This is related to the earlier observations that the TCCs for option 3 are slightly but consistently higher than those for option 1 on the low and high ends of the SS range. This means that the same low percentage of maximum possible number-correct score would always lead to a higher SS under option 1 than under option 3 at low and high ends. The lack of tau-equivalence on the high end is visible only in the grade 3 plots, simply because no students are present in the range at the other two grades.

CR context

The RS-SS tables for the CR context are provided in Tables 9 – 11, respectively, for grades 3, 5, and 8. In each table, the RS-SS pairs are listed for options 1 – 3, from left to right. The column "Diff. SS" is absent from Tables 9 and 11, because the maximum number of raw-score points for the CR2 form does not equal twice the number of raw-score points for the CR1 form, which necessitated the computation of the percentage of the maximum raw-score points (% NC) for each of the CR2, weighted CR1, and unweighted CR1 forms.

Comparisons of the RS-SS tables and the test characteristic curves (TCCs)
The RS-SS pairs for the CR context are graphically depicted in Figures 13 – 15 as test characteristic curves (TCCs). In Figure 15, the option 1 and option 2 TCCs appear on top of each other, although they, in fact, differ by up to 10 SS points as indicated in Table 11.

As in the ER context, the option 2 TCC is nearly identical or very comparable to the option 1 TCC. The option 3 TCC is still substantially comparable to the criterion option 1 TCC through the range of SSs, although they are somewhat apart at the lower asymptote in grades 3 and 5, that is, in a SS range below 350, and, as seen in the ER context, consistently higher than the target TCC in the high and low ends of the SS range.

Although the number of ER and CR items being doubly weighted increases over three grades (3, 5, and 7 items for grades 3, 5, and 8), no systematic differences between options 1 and 2 in terms of the TCCs are observed over the grades.

Comparisons in standard errors (SEs) and SE curves
The constrained standard errors (SEs) for the SSs in Tables 9 – 11 are plotted in Figures 16 - 18. The RS-SS tables and the SE plots for the CR context manifest similar patterns that were seen in the ER context. Namely,

- in all cases, the SE curve for option 1 (CR2), as expected, is the lowest throughout the SS range,
- as far as the low end of the SS range below 350 is concerned, the SE curve for option 2 (weighted CR1) is closer to the criterion SE curve for option 1 (unweighted CR2) than is the SE curve for option 3 (unweighted CR1). The option 3 SE curve deviates somewhat more from the option 1 SE curve in the CR context than in the ER context. Namely, as compared with the ER context, the SEs for option 3 in the CR context are higher than those of option 1 by even more (i.e., by about 15 to 30 SS points) at the lower end,
- for the mid-range of SSs between 350 and 600, the three SE curves for options 1 – 3 are still very comparable. Although the option 1 SE curve in the CR context is discernibly the lowest, the differences of SEs between option 1 and option 2 and between option 1 and option 3 are very small and largely within 5 SS points.
- on the high end of the SS range above 600, although the option 2 curve tends to show slightly higher SE than the option 3 curve, they are once again very similar in terms of SEs. The SE plot for either option is substantially greater than that for the criterion SE for option 1. At a SS of 700, the linearly-interpolated SE for either option 2 or option 3 is about 13 to 25 SS points higher than that for the option 1 SE.

The locations of the weighted items range between 383 and 473 for grade 3, between 377 and 525 for grade 5, and between 350 and 506. These locations seem to explain why the weighted CR1 SE curves are lower than the unweighted CR1 curves in the lower ends of the SS range.

Scale score comparisions
As in the ER context, for each grade, three sets of scale scores (SSs) for the students in the scaling sample were compared; that is, option 1 SSs, option 2 SSs, and option 3 SSs. The pairs of SSs for the CR context are plotted in Figures 19 through 24. At each grade, the option 1 SSs

are plotted against option 2 SSs on the left, and then against the option 3 SSs on the right. The plots also include the correlations. As in the ER context, the correlations between the option 1 and option 2 SSs are very similar to those between the option 1 and option 3 SSs.

Previously in the ER context, it was observed that the option 1 – option 3 SS pairs hugged the 45-degree line more closely than the option 1 – option 2 SS pairs. This observation is much less apparent in the plots for the CR context. However, the lack of tau-equivalence at the low end of the SS range for the option 1 – option 3 scatterplots is still visible in the CR context. As noted earlier, this is in line with what has been observed in terms of TCC curves. As in the ER context, the option 1 – option 2 scatterplots for the CR context appear tau-equivalent.

The numerical results of the scale-score comparisons in the CR context are summarized in Table 13. As in the ER context (Table 12), both the option 2 and option 3 SSs are, on average, very comparable to the criterion option 1 SSs in all three grades. The differences in the SSs between option 1 and option 2 are, on average, within a half SS point, while they are approximately within a SS point between option 1 and option 3. The SS standard deviations (SDs) are very comparable among the three options, although the option 2 and option 3 SDs are more similar to each other than are the option 1and option 2 SDs, or the option 1 and option 3 SDs.

The absolute mean difference score for the option 1 – option 3 comparison tends to increase from grade 3 to grade5 to grade 8 (|-.25|; |.54|; |1.12|). This may be caused by the increasing number of additional ER and CR items that the CR2 forms contain (3, 5, and 7 items in grades 3, 5, and 8). That is, the difference in test length and in items between the option 1 (CR2) and option 3 (CR1) forms increases from the lowest grade to the highest grade. However, Table 12 for the ER context displays a similar pattern of quasi-increasing mean difference scores over grades, even though the difference between the ER2 and ER1 forms remains constant across grades (i.e., one ER item). Therefore, the increase over grades in the absolute mean difference score in the CR context may well be a coincidence.

As before, the percentage of students with relatively small within-student SS differences between option 1 and option 2 was compared with that for the option 1 - option 3 comparison. As before, "relatively small" was defined using two cut-off differences: 5 SS points or less, and 10 SS points or less, in absolute value. Regardless of the grades, the weighting option has produced scores that are within 5 SS points in 30% - 35% of the students and within 10 SS points in 53% - 63%. The option of no weighting has generated scores that are within 5 SS points in 33 % - 42%, and within 10 SS points in 59% - 64% of the students. Irrespective of the cut-off differences, the option 3 SSs are, consistently in all three grades, similar to the option 1 SSs in 1% - 9% more students than are the option 2 SSs. A similar observation was made in the ER context.

Discussion

The weighting option, option 2, generally has produced percentage-raw-score-to-scale-score conversion tables that are more similar to the criterion tables (option 1) than the no-weighting option (option 3) throughout the scale-score range from 200 to 800. The option-3 tables slightly

differ from the target tables on the extreme ends. In terms of measurement error associated with scale scores, once again, option 2 has slightly or substantially lower standard errors than option 3, except on the high end where option 3 has slightly lower errors. Smaller errors for the weighted scale scores in the lower range reflect the fact that the items weighted twice happen to be located in this range. In a way, the lower SEs for option 2 are expected, since some items are considered twice in the computation of the errors even though they were never actually taken twice by the students. It would be interesting to see if lower SEs for option 2 could be validated in a test-retest or alternate-form study.

These relative similarities between the criterion option 1 and the weighting option with regard to the conversion tables do not translate straight into student scores. The three options are almost identical in terms of average scale scores. This is not surprising in light of the result of an unreported analysis that 92% - 96% of scores are located in a middle range between 350 and 600 where the three options are substantially similar. Scatterplots of scale scores have revealed that in terms of marginal mean scale scores, the weighted option-2 mean scores seem to be more similar to the criterion mean scores than do the unweighted option-3 mean scores at the low and high ends of SS range. Thus, the weighting option compares very favorably against the no-weighting option in terms of both overall and marginal mean scores.

The comparisons of options 2 and 3 relative to option 1 at the level of individual student scores have presented a slightly different picture. In the ER context, unweighted scores (option 3) are within 10 scale-score points of their criterion option-1 scores in 2% - 5% more students than are weighted scores (option 2), and within 5 points in 14% - 15% more students in two of the three grades. In the CR context, unweighted scores are within 10 points of the target scores in 1% - 9% more students than are weighted scores and within 5 points in 3% - 7% more students. Thus, the no-weighting option (option 3) tends to achieve closer approximation to the criterion scores than does the weighting option (option 2) at the level of individual students. However, if |10| scale-score deviations are acceptable, the 2% - 5% increases in less accurate scores with the weighting option in the ER context do not appear detrimental. Ten-point differences may be endured particularly in a situation where students are classified into categories, and their exact scores are not as crucial. Unfortunately, the state whose data were used for the study has not established performance cut-off scores, and the study could not evaluate the options in terms of their effects on student classifications into performance categories. The use of a shorter form and of explicit weighting may have an even smaller impact on performance classifications, particularly in the ER context. Even the 1% - 9% decreases in score precision in the CR context may be within a threshold under some circumstances.

Although the evaluation of scale-score differences used two criteria (i.e., 5 and 10 points in absolute magnitude), the |5|-point criterion may seem too stringent. For example, differences seen between pairs of scale scores from number-correct (NC) and pattern scoring methods can be greater than |5| points. A separate analysis of data from another state based on two tests containing mixed item types has demonstrated that most (98% or 99%) of NC scores were within 10 points of their corresponding pattern scores, while 87% - 90% of NC scores were within 5 points. Despite documented increases in accuracy of scores by pattern scoring (e.g., Yen & Candell, 1991), many practitioners opt for NC scoring for the reason that it is easier to understand, meaning that the magnitude of a decrease in precision associated with NC scoring is

well tolerated or even accepted. Thus, the |10|-point criterion, as opposed to the |5|-point criterion, may be emphasized in assessing options 2 and 3 against option 1.

The discussion so far has focused on the impact of deliberate weighting. How about the impact of test length? The effect of test length on score accuracy can be evaluated by comparing scores between option 1 and option 3. Options 1 and 3 differ in only one feature, test length. As expected, the effect of test length is considerably greater in the CR context than in the ER context. The difference between the longer and shorter forms in the ER context is a single ER item, whereas it is a set of CR and ER items in the CR context. Irrespective of the |5|- or |10|-point criterion, the percentages of students with unweighted scores that are relatively similar to the criterion option-1 scores range in the 60% - 90%s for the ER context. The percentages diminish markedly in the CR context, between about 35% and 65%. Note that these appreciable drops in score precision in the CR context are caused by lacking only several (3 – 7 ER and CR) additional items.

Weighting or no weighting, some practitioners would expect a higher level of score accuracy and favor a longer test, while others would tolerate lower precision and accept a shorter test. There are several factors that would impact a decision as to which way to go. For example, an obvious factor is the level of stakes involved in the decisions to be made based on student scores. Second, one should consider whether it is a single score, or a category in which a student is placed, that is crucial. Third, as noted in the Introduction, testing time may be a practical consideration. Fourth, as seen in Table 1, the tests with a single ER or a single set of CR/ER items with no weighting have reliabilities very similar to those with twice as many ER and/or CR items. Namely, in terms of test reliability, the shorter form should be considered acceptable. Some of the same factors may play a role in deciding between number-correct and item-pattern scoring.

The study has other limitations. The study treated student scores from a longer form as the criterion scores. However, these scores may still differ from the *true* scores, and a simulation is called for. Moreover, the study did not address the aspect of content representation of the different forms. The form with a single set of ER and CR items with no weighting obviously has different content coverage than the forms with twice as many of these items or the "virtual" forms when these items are weighted twice. This could be a serious drawback for the no-weighting option. Furthermore, the live-data study has a few features that may not be shared by other assessments, such as the way the ER items were scored, and the fact that the doubly-weighted items were relatively easy. Due to these idiosyncrasies, the results, to some extent, may not be generalizable to other tests.

In conclusion, the desire to explicitly "honor" the efforts taken by students to produce longer responses is understandable. This study has found no loss in the precision of average scores, both overall and throughout the range of scale scores, due to intentional double-weighting. On the level of individual student scores, deliberate weighting, as compared with no weighting, results in more students with less accurate scores. The increase in less accurate scores is relatively small – 5% on average – if differences from the criterion scores from a longer test up

to ten points are acceptable. However, it is clear that purposeful weighting can never overcome the loss of score accuracy caused by shorter test length.[2]

[2] Options for future research include (1) a simulation study, and (2) another alternative of summing two ratings for each CR item if each response is rated by two raters, thereby doubling the number of score points contributed by each CR item, which is a different way of accomplishing the desired double-weighting.

References

Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika, 37*, 29-51.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: An application of an EM algorithm. *Psychometrika, 46*, 443-459.

Burket, G. R. (1991). *PARDUX* [Computer program]. Unpublished.

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

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Menlo Park, CA: Addison-Wesley.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. Applied Psychological Measurement, 16, 159–176.

Muraki, E., & Bock, R. D. (1991). *PARSCALE: Parameter Scaling of Rating Data* [Computer program]. Chicago, IL: Scientific Software, Inc.

Thissen, D. (1982). Marginal maximum likelihood estimation for the one-parameter logistic model. *Psychometrika, 47*, 175-186.

Thissen, D. (1991). *MULTILOG* [Computer program]. Chicago, IL: Scientific Software, Inc.

Wright, B. D., & Linacre, J. M. (1992). *BIGSTEPS Rasch Analysis* [Computer program]. Chicago, IL: MESA Press.

Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement, 30*, 187-213.

Yen, W. M., & Candell, G. L. (1991). Increasing score reliability with item-pattern scoring: An empirical study in five score metrics. *Applied Measurement in Education, 4*, 209-228.

Appendix A : Item response theory models used in the study

Because the characteristics of selected-response (SR) and constructed-response (CR) items are different, two IRT models were used in item calibration. The three-parameter logistic model (3PL) (Lord & Novick, 1968; Lord, 1980) was used in the analysis of SR items. In this model, the probability that a student with ability θ responds correctly to item i is

$$P_i(\theta) = c_i + \frac{1 - c_i}{1 + \exp[-1.7a_i(\theta - b_i)]},$$

where a_i is the item discrimination, b_i is the item difficulty, and c_i is the probability of a correct response by a very low-scoring student.

For analysis of the constructed-response items, the two-parameter partial credit model (2PPC) (Muraki, 1992; Yen, 1993) was used. The 2PPC model is a special case of Bock's (1972) nominal model. Bock's model states that the probability of an examinee with ability θ having a score (k-1) at the k-th level of the j-th item is

$$P_{jk}(\theta) = P(x_j = k-1 \mid \theta) = \frac{\exp Z_{jk}}{\sum_{i=1}^{m_j} \exp Z_{ji}}, \quad k = 1 \ldots m_j,$$

where

$$Z_{jk} = A_{jk}\theta + C_{jk}.$$

The m_j denotes the number of score levels for the j-th item, and typically, the highest score level is assigned $(m_j - 1)$ score points. For the special case of the 2PPC model used here, the following constraints were used:

$$A_{jk} = \alpha_j(k-1),$$

and

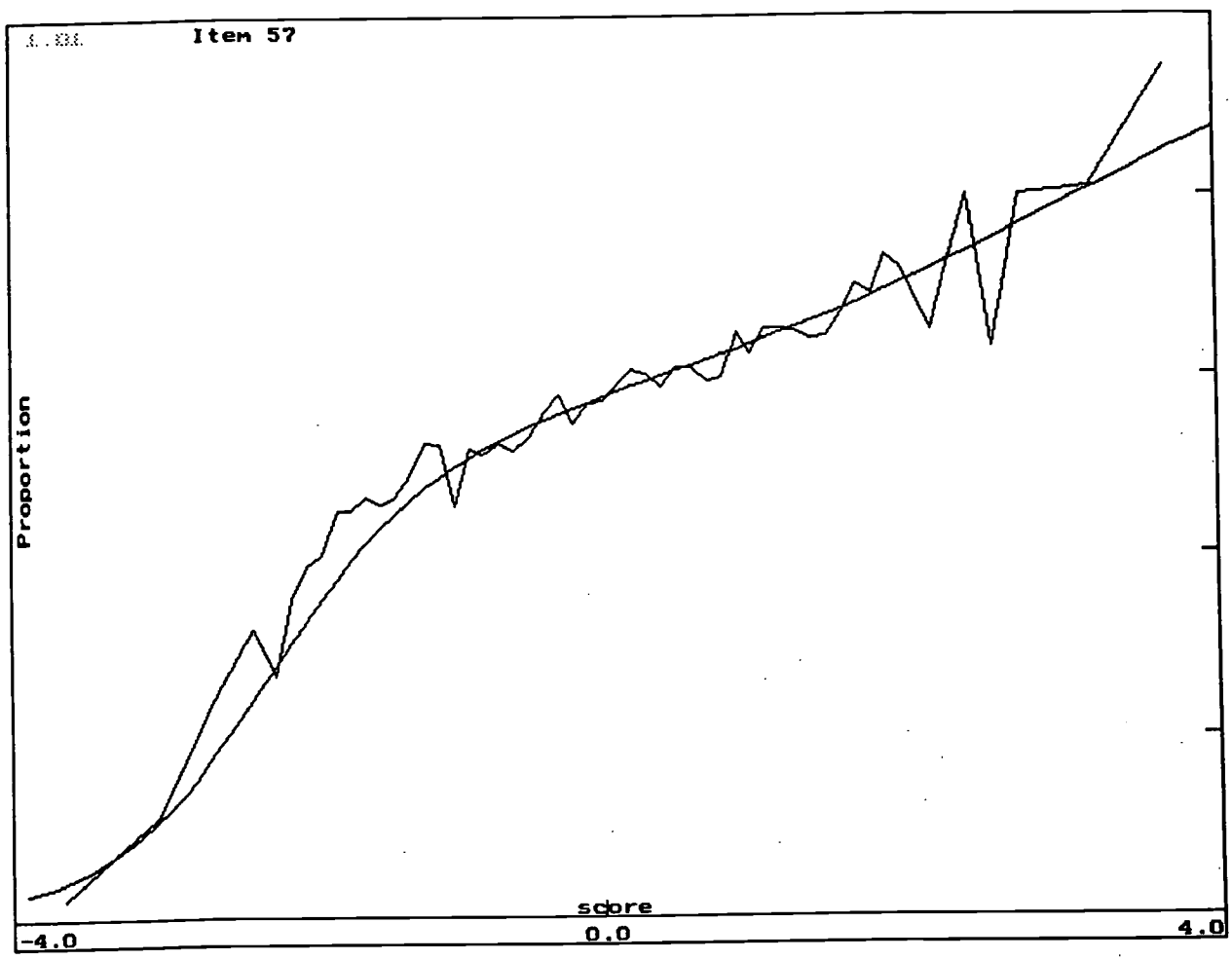
$$C_{jk} = -\sum_{i=0}^{k-1} \gamma_{ji}, \text{ where } \gamma_{j0} = 0,$$

where α_j and γ_{ji} are the free parameters to be estimated from the data. Each item has $(m_j - 1)$ independent γ_{ji} parameters and one α_j parameter; a total of m_j parameters are estimated for each item.

The IRT model parameters were estimated using CTB's PARDUX software (Burket, 1991). PARDUX estimates parameters simultaneously for SR and CR items using marginal maximum likelihood procedures implemented via the EM (expectation-maximization) algorithm (Bock & Aitkin, 1981; Thissen, 1982).

Simulation studies have compared PARDUX with MULTILOG (Thissen, 1991), PARSCALE (Muraki & Bock, 1991), and BIGSTEPS (Wright & Linacre, 1992). PARSCALE, MULTILOG, and BIGSTEPS are among the most widely known and used IRT programs. PARDUX was found to perform at least as well as these other programs.

Appendix B



List of Tables and Figures

Table 1.
Grade 3 Mean P-Values and Reliabilities (N = 2,574)

Item Type	Item #	P-value	ER2	ER1	CR2	CR1
SR	1	.520	√	√	√	√
	2	.640	√	√	√	√
	3	.780	√	√	√	√
	4	.619	√	√	√	√
	5	.371	√	√	√	√
	6	.638	√	√	√	√
	7	.805	√	√	√	√
	8	.787	√	√	√	√
	9	.723	√	√	√	√
	10	.768	√	√	√	√
	11	.817	√	√	√	√
	12	.788	√	√	√	√
	13	.804	√	√	√	√
	14	.533	√	√	√	√
	15	.636	√	√	√	√
	16	.655	√	√	√	√
	17	.760	√	√	√	√
	18	.669	√	√	√	√
	19	.689	√	√	√	√
	20	.749	√	√	√	√
	21	.712	√	√	√	√
	22	.785	√	√	√	√
	23	.850	√	√	√	√
	24	.840	√	√	√	√
	25	.408	√	√	√	√
	26	.564	√	√	√	√
	27	.838	√	√	√	√
	28	.785	√	√	√	√
	29	.924	√	√	√	√
CR	1	.681			√	
	2	.606	√	√	√	√
	3	.781	√	√	√	√
	4	.751			√	
	5	.740	√	√		
ER	1	.598	√		√	
	2	.576	√	√	√	√
Mean p-value			.699	.702	.699	.701
Feldt-Raju reliability			.911	.907	.917	.905

Mean p of ER/CRs that are in CR2 but not in CR1 = .68
Mean p of ER/CRs that are doubly weighted = .65

Table 2.
Grade 5 Mean P-Values and Reliabilities (N = 2,642)

Item Type	Item #	P-value	ER2	ER1	CR2	CR1
SR	1	.500	√	√	√	√
	2	.785	√	√	√	√
	3	.613	√	√	√	√
	4	.761	√	√	√	√
	5	.697	√	√	√	√
	6	.568	√	√	√	√
	7	.311	√	√	√	√
	8	.265	√	√	√	√
	9	.453	√	√	√	√
	10	.783	√	√	√	√
	11	.776	√	√	√	√
	12	.798	√	√	√	√
	13	.485	√	√	√	√
	14	.741	√	√	√	√
	15	.690	√	√	√	√
	16	.757	√	√	√	√
	17	.282	√	√	√	√
	18	.902	√	√	√	√
	19	.766	√	√	√	√
	20	.575	√	√	√	√
	21	.594	√	√	√	√
	22	.488	√	√	√	√
CR	1	.712	√	√	√	√
	2	.684	√	√	√	√
	3	.702	√	√	√	
	4	.738	√	√	√	
	5	.627	√	√	√	
	6	.439	√	√	√	√
	7	.559	√	√	√	
	8	.841			√	√
ER	1	.606	√	√	√	√
	2	.556	√		√	
Mean p-value			.620	.622	.627	.625
Feldt-Raju reliability			.879	.873	.885	.848

Mean p of ER/CRs that are in CR2 but not in CR1 = .64
Mean p of ER/CRs that are doubly weighted = .66

Table 3.
Grade 8 Mean P-Values and Reliabilities (N = 3,633)

Item Type	Item #	P-value	ER2	ER1	CR2	CR1
SR	1	.811	√	√	√	√
	2	.789	√	√	√	√
	3	.624	√	√	√	√
	4	.823	√	√	√	√
	5	.836	√	√	√	√
	6	.622	√	√	√	√
	7	.835	√	√	√	√
	8	.595	√	√	√	√
	9	.679	√	√	√	√
	10	.668	√	√	√	√
	11	.826	√	√	√	√
	12	.370	√	√	√	√
	13	.378	√	√	√	√
	14	.692	√	√	√	√
	15	.612	√	√	√	√
	16	.433	√	√	√	√
	17	.731	√	√	√	√
	18	.927	√	√	√	√
	19	.317	√	√	√	√
	20	.722	√	√	√	√
	21	.517	√	√	√	√
	22	.652	√	√	√	√
	23	.451	√	√	√	√
	24	.461	√	√	√	√
	25	.439	√	√	√	√
	26	.763	√	√	√	√
CR	1	.599	√	√	√	√
	2	.707	√	√	√	
	3	.283	√	√		
	4	.375	√	√	√	√
	5	.495	√	√	√	√
	6	.818	√	√	√	√
	7	.927			√	
	8	.940			√	√
	9	.741			√	
	10	.456			√	
	11	.560			√	√
	12	.378			√	
	13	.753			√	
ER	1	.657	√	√	√	√
	2	.640	√		√	
Mean p-value			.622	.621	.640	.637
Feldt-Raju reliability			.889	.883	.914	.889

Mean p of ER/CRs that are in CR2 but not in CR1 = .66
Mean p of ER/CRs that are doubly weighted = .64

Table 4.
Numbers of SR, CR, ER Items and Score Points, and Objective Coverage

	Form			
	ER2	ER1	CR2	CR1
	Grade 3			
# SR items	29	29	29	29
# CR items	3	3	4	2
# 2-pt items	3	3	3	2
# 3-pt items			1	
# ER items (5 pts each)	2	1	2	1
Total # items	34	33	35	32
Total # score points	45	40	48	38
# of CR / ER items in:				
Obj. 3	2	1	2	1
Obj. 4	1	1	2	1
Obj. 5	1	1		
Obj. 6	1	1	2	1
	Grade 5			
# SR items	22	22	22	22
# CR items (2 pts each)	7	7	8	4
# ER items (5 pts each)	2	1	2	1
Total # items	31	30	32	27
Total # score points	46	41	48	35
# of CR / ER items in:				
Obj. 2	4	4	4	2
Obj. 5	2	1	2	1
Obj. 6	3	3	4	2
	Grade 8			
# SR items	26	26	26	26
# CR items	6	6	12	6
# 2-pt items	6	6	11	5
# 4-pt items			1	1
# ER items (5 pts each)	2	1	2	1
Total # items	34	33	40	33
Total # score points	48	43	62	45
# of CR / ER items in:				
Obj. 2	2	2	4	2
Obj. 3	2	2	5	2
Obj. 4	1	1	2	1
Obj. 5	1	1	1	1
Obj. 6	2	1	2	1

Table 5.
Option - Context Combinations and Where Intentional (Un)Weighting was Used

ER Context

Option 1 (ER2)	Option 2 (Weighted ER1)	Option 3 (Unweighted ER1)
	Grade 3	
2 ERs 3 CRs 29 MCs	1 ER weighted 3 CRs 29 MCs	1 ER unweighted 3 CRs 29 MCs
	Grade 5	
2 ERs 7 CRs 22 MCs	1 ER weighted 7 CRs 22 MCs	1 ER unweighted 7 CRs 22 MCs
	Grade 8	
2 ERs 6 CRs 26 MCs	1 ER weighted 6 CRs 26 MCs	1 ER unweighted 6 CRs 26 MCs

CR Context

Option 1 (CR2)	Option 2 (Weighted CR1)	Option 3 (Unweighted CR1)
	Grade 3	
2 ERs 4 CRs 29 MCs	1 ER weighted 2 CRs weighted 29 MCs	1 ER unweighted 2 CRs unweighted 29 MCs
	Grade 5	
2 ERs 8 CRs 22 MCs	1 ER weighted 4 CRs weighted 22 MCs	1 ER unweighted 4 CRs unweighted 22 MCs
	Grade 8	
2 ERs 12 CRs 26 MCs	1 ER weighted 6 CRs weighted 26 MCs	1 ER unweighted 6 CRs unweighted 26 MCs

Table 6.
Grade 3 RS-to-SS Tables : Op. 1 (ER2), Op. 2 (Weighted ER1), and Op.3 (Unweighted ER1)

		Option 1 : ER2		Op. 2 : Weighted ER1		Diff. SS	Op. 3 : Unweighted ER1			
RS	% NC	SS	SE	SS	SE	(Op1-Op2)	RS	% NC	SS	SE
0	0	200	198	200	200	0	0	0	200	210
1	2	200	198	200	200	0	1	3	200	210
2	4	200	198	200	200	0	2	5	200	210
3	7	200	198	200	200	0	3	8	200	210
4	9	200	198	200	200	0	4	10	200	210
5	11	200	198	200	200	0	5	13	200	210
6	13	200	198	200	200	0	6	15	200	210
7	16	200	198	200	200	0	7	18	200	210
8	18	314	84	311	89	3	8	20	321	89
9	20	342	56	340	61	2	9	23	351	59
10	22	359	41	356	46	3	10	25	369	44
11	24	371	34	369	38	2	11	28	383	36
12	27	381	30	379	34	2	12	30	394	32
13	29	390	27	389	31	1	13	33	404	28
14	31	399	25	398	29	1	14	35	413	26
15	33	406	24	406	27	0	15	38	422	24
16	36	413	23	413	25	0	16	40	429	22
17	38	420	22	420	24	0	17	43	436	20
18	40	427	21	427	22	0	18	45	443	19
19	42	433	20	434	21	-1	19	48	449	18
20	44	439	19	440	20	-1	20	50	455	17
21	47	445	18	446	19	-1	21	53	461	17
22	49	451	17	452	18	-1	22	55	466	16
23	51	456	17	457	17	-1	23	58	472	16
24	53	462	16	462	17	0	24	60	477	15
25	56	467	16	468	16	-1	25	63	483	15
26	58	472	16	473	16	-1	26	65	488	15
27	60	478	15	478	16	0	27	68	494	15
28	62	483	15	484	16	-1	28	70	499	15
29	64	488	15	489	15	-1	29	73	505	15
30	67	494	15	494	15	0	30	75	511	16

(continued)

Table 6.
Grade 3 RS-to-SS Tables : Op. 1 (ER2), Op. 2 (Weighted ER1), and Op.3 (Unweighted ER1)

		Option 1 : ER2		Op. 2 : Weighted ER1		Diff. SS	Op. 3 : Unweighted ER1			
RS	% NC	SS	SE	SS	SE	(Op1-Op2)	RS	% NC	SS	SE
31	69	499	15	500	15	-1	31	78	518	16
32	71	505	15	505	16	0	32	80	526	17
33	73	511	15	512	16	-1	33	83	534	19
34	76	517	16	518	17	-1	34	85	544	21
35	78	524	17	525	18	-1	35	88	556	23
36	80	532	18	533	19	-1	36	90	571	26
37	82	541	20	543	21	-2	37	93	591	31
38	84	552	21	553	23	-1	38	95	621	43
39	87	564	23	566	26	-2	39	98	681	66
40	89	579	26	582	31	-3	40	100	800	135
41	91	599	32	602	38	-3				
42	93	627	40	630	50	-3				
43	96	668	50	670	64	-2				
44	98	728	63	726	82	2				
45	100	800	97	800	134	0				

Table 7.
Grade 5 RS-to-SS Tables : Op. 1 (ER2), Op. 2 (Weighted ER1), and Op. 3 (Unweighted ER1)

		Option 1 : ER2		Op. 2 : Weighted ER1		Diff. SS	Op. 3 : Unweighted ER1			
RS	% NC	SS	SE	SS	SE	(Op1-Op2)	RS	% NC	SS	SE
0	0	200	185	200	168	0	0	0	200	192
1	2	200	185	200	168	0	1	2	200	192
2	4	200	185	200	168	0	2	5	200	192
3	7	200	185	200	168	0	3	7	200	192
4	9	200	185	200	168	0	4	10	200	192
5	11	200	185	247	121	-47	5	12	200	192
6	13	258	126	293	75	-35	6	15	260	132
7	15	307	77	320	55	-13	7	17	311	82
8	17	333	51	339	44	-6	8	20	338	55
9	20	350	39	354	37	-4	9	22	357	43
10	22	363	32	366	33	-3	10	24	372	36
11	24	374	28	376	30	-2	11	27	384	32
12	26	383	26	385	28	-2	12	29	394	29
13	28	391	25	393	27	-2	13	32	404	28
14	30	399	24	401	26	-2	14	34	413	26
15	33	406	23	408	26	-2	15	37	422	25
16	35	414	23	415	25	-1	16	39	431	24
17	37	421	23	423	25	-2	17	41	439	24
18	39	428	23	430	24	-2	18	44	447	23
19	41	436	22	436	24	0	19	46	454	22
20	43	443	22	443	24	0	20	49	462	22
21	46	450	22	450	23	0	21	51	469	21
22	48	457	21	457	23	0	22	54	476	21
23	50	464	21	463	22	1	23	56	483	20
24	52	471	20	470	22	1	24	59	490	20
25	54	478	20	476	22	2	25	61	497	20
26	57	485	20	483	22	2	26	63	504	19
27	59	491	20	489	22	2	27	66	512	19
28	61	498	19	496	21	2	28	68	519	19
29	63	505	19	502	21	3	29	71	526	19
30	65	512	19	509	21	3	30	73	534	20

(continued)

Table 7.
Grade 5 RS-to-SS Tables : Op. 1 (ER2), Op. 2 (Weighted ER1), and Op. 3 (Unweighted ER1)

		Option 1 : ER2		Op. 2 : Weighted ER1		Diff. SS	Op. 3 : Unweighted ER1			
RS	% NC	SS	SE	SS	SE	(Op1-Op2)	RS	% NC	SS	SE
31	67	519	19	515	21	4	31	76	542	20
32	70	527	19	522	21	5	32	78	550	20
33	72	534	19	529	22	5	33	80	559	21
34	74	542	20	537	22	5	34	83	569	22
35	76	550	20	544	22	6	35	85	580	23
36	78	559	21	553	23	6	36	88	593	25
37	80	568	22	561	24	7	37	90	608	28
38	83	579	23	571	25	8	38	93	626	32
39	85	591	24	581	26	10	39	95	650	38
40	87	604	26	592	28	12	40	98	690	53
41	89	620	28	606	30	14	41	100	800	151
42	91	639	32	621	34	18				
43	93	662	36	640	38	22				
44	96	692	41	665	46	27				
45	98	737	54	705	65	32				
46	100	800	90	800	159	0				

Table 8.
Grade 8 RS-to-SS Tables : Op. 1 (ER2), Op. 2 (Weighted ER1), and Op. 3 (Unweighted ER1)

		Option 1 : ER2		Op. 2 : Weighted ER1		Diff. SS	Op. 3 : Unweighted ER1			
RS	% NC	SS	SE	SS	SE	(Op1-Op2)	RS	% NC	SS	SE
0	0	200	180	200	187	0	0	0	200	194
1	2	200	180	200	187	0	1	2	200	194
2	4	200	180	200	187	0	2	5	200	194
3	6	200	180	200	187	0	3	7	200	194
4	8	200	180	200	187	0	4	9	200	194
5	10	200	180	200	187	0	5	12	200	194
6	13	272	108	273	114	-1	6	14	274	120
7	15	318	62	320	67	-2	7	16	323	71
8	17	339	41	342	45	-3	8	19	347	47
9	19	352	31	355	34	-3	9	21	362	35
10	21	362	26	365	29	-3	10	23	373	30
11	23	370	24	373	26	-3	11	26	383	27
12	25	377	22	380	25	-3	12	28	392	26
13	27	385	22	387	24	-2	13	30	401	26
14	29	392	22	393	24	-1	14	33	410	25
15	31	399	23	400	24	-1	15	35	418	25
16	33	406	23	407	24	-1	16	37	426	24
17	35	414	23	414	24	0	17	40	434	23
18	38	421	23	421	23	0	18	42	441	22
19	40	428	22	429	23	-1	19	44	448	21
20	42	436	21	436	22	0	20	47	455	21
21	44	442	21	442	21	0	21	49	462	20
22	46	449	20	449	21	0	22	51	469	20
23	48	456	20	456	20	0	23	53	476	19
24	50	462	20	462	20	0	24	56	482	19
25	52	469	19	469	20	0	25	58	489	19
26	54	475	19	475	19	0	26	60	496	19
27	56	482	19	481	19	1	27	63	503	19
28	58	488	19	488	19	0	28	65	510	19
29	60	495	19	494	19	1	29	67	517	20
30	63	501	19	500	19	1	30	70	524	20

(continued)

Table 8.
Grade 8 RS-to-SS Tables : Op. 1 (ER2), Op. 2 (Weighted ER1), and Op. 3 (Unweighted ER1)

		Option 1 : ER2		Op. 2 : Weighted ER1		Diff. SS	Op. 3 : Unweighted ER1			
RS	% NC	SS	SE	SS	SE	(Op1-Op2)	RS	% NC	SS	SE
31	65	508	19	507	19	1	31	72	532	20
32	67	515	19	514	19	1	32	74	540	21
33	69	522	19	521	20	1	33	77	549	22
34	71	529	19	528	20	1	34	79	559	23
35	73	536	20	535	21	1	35	81	570	25
36	75	544	20	543	21	1	36	84	582	27
37	77	553	21	552	22	1	37	86	597	30
38	79	562	23	561	23	1	38	88	614	34
39	81	573	24	571	25	2	39	91	636	38
40	83	585	26	583	27	2	40	93	663	44
41	85	598	28	596	30	2	41	95	701	56
42	88	614	31	612	33	2	42	98	781	110
43	90	633	34	630	36	3	43	100	800	126
44	92	655	37	651	39	4				
45	94	682	41	676	45	6				
46	96	717	52	711	58	6				
47	98	788	100	785	108	3				
48	100	800	110	800	121	0				

Table 9.
Grade 3 RS-to-SS Tables : Op. 1 (CR2), Op. 2 (Weighted CR1) and Op. 3 (Unweighted CR1)

	Option 1 : CR2			Op. 2 : Weighted CR1			Op. 3 : Unweighted CR1		
RS	% NC	SS	SE	% NC	SS	SE	% NC	SS	SE
0	0	200	196	0	200	200	0	200	214
1	2	200	196	2	200	200	3	200	214
2	4	200	196	4	200	200	5	200	214
3	6	200	196	6	200	200	8	200	214
4	8	200	196	9	200	200	11	200	214
5	10	200	196	11	200	200	13	200	214
6	13	200	196	13	200	200	16	200	214
7	15	200	196	15	200	200	18	200	214
8	17	308	88	17	300	100	21	328	85
9	19	338	58	19	332	69	24	357	57
10	21	355	42	21	350	51	26	375	43
11	23	368	34	23	363	42	29	389	36
12	25	378	30	26	374	37	32	401	31
13	27	387	27	28	383	34	34	411	28
14	29	395	25	30	392	32	37	420	25
15	31	402	24	32	400	30	39	428	23
16	33	409	22	34	407	28	42	436	21
17	35	416	21	36	415	27	45	443	19
18	38	422	20	38	421	25	47	449	18
19	40	428	19	40	428	24	50	456	18
20	42	434	19	43	434	22	53	462	17
21	44	439	18	45	440	21	55	467	16
22	46	444	17	47	446	20	58	473	16
23	48	449	17	49	451	20	61	479	15
24	50	454	16	51	457	19	63	485	15
25	52	459	16	53	462	18	66	490	15
26	54	464	15	55	467	18	68	496	15
27	56	469	15	57	472	18	71	502	15
28	58	474	15	60	477	17	74	509	15
29	60	478	15	62	483	17	76	516	16
30	63	483	15	64	488	17	79	523	17

(continued)

Table 9.
Grade 3 RS-to-SS Tables : Op. 1 (CR2), Op. 2 (Weighted CR1) and Op. 3 (Unweighted CR1)

	Option 1 : CR2			Op. 2 : Weighted CR1			Op. 3 : Unweighted CR1		
RS	% NC	SS	SE	% NC	SS	SE	% NC	SS	SE
31	65	488	14	66	493	17	82	532	18
32	67	493	14	68	498	17	84	542	20
33	69	498	15	70	504	17	87	554	22
34	71	503	15	72	509	18	89	569	25
35	73	508	15	74	516	18	92	588	30
36	75	514	16	77	522	19	95	618	42
37	77	521	16	79	530	21	97	678	66
38	79	528	17	81	538	23	100	800	135
39	81	535	18	83	548	25			
40	83	544	20	85	559	27			
41	85	555	22	87	572	30			
42	88	567	24	89	588	35			
43	90	582	27	91	608	44			
44	92	602	32	94	636	55			
45	94	629	40	96	675	68			
46	96	670	50	98	730	85			
47	98	729	63	100	800	136			
48	100	800	97						

Table 10.
Grade 5 RS-to-SS Tables : Op. 1 (CR2), Op. 2 (Weighted CR1), and Op. 3 (Unweighted CR1)

		Option 1 : CR2		Op. 2 : Weighted CR1		Diff. SS	Op. 3 : Unweighted CR1			
RS	% NC	SS	SE	SS	SE	(Op1-Op2)	RS	% NC	SS	SE
0	0	200	176	200	181	0	0	0	200	203
1	2	200	176	200	181	0	1	3	200	203
2	4	200	176	200	181	0	2	6	200	203
3	6	200	176	200	181	0	3	9	200	203
4	8	200	176	200	181	0	4	11	200	203
5	10	200	176	200	181	0	5	14	200	203
6	13	238	138	222	158	16	6	17	269	134
7	15	290	85	278	103	12	7	20	321	82
8	17	319	57	309	73	10	8	23	349	55
9	19	339	43	330	57	9	9	26	367	42
10	21	353	35	345	47	8	10	29	382	36
11	23	364	30	358	40	6	11	31	395	33
12	25	374	27	368	36	6	12	34	406	31
13	27	382	25	377	33	5	13	37	417	29
14	29	390	24	386	31	4	14	40	428	28
15	31	398	23	393	30	5	15	43	438	27
16	33	405	23	401	29	4	16	46	448	26
17	35	412	23	408	29	4	17	49	457	25
18	38	419	22	415	28	4	18	51	466	24
19	40	426	22	422	28	4	19	54	475	23
20	42	433	22	429	28	4	20	57	484	23
21	44	440	22	436	27	4	21	60	493	22
22	46	447	21	443	27	4	22	63	502	22
23	48	453	21	450	27	3	23	66	510	21
24	50	460	21	457	26	3	24	69	519	21
25	52	467	20	464	26	3	25	71	528	21
26	54	474	20	471	25	3	26	74	538	21
27	56	480	20	477	25	3	27	77	548	22
28	58	487	20	484	25	3	28	80	558	22
29	60	494	19	491	24	3	29	83	570	23
30	63	500	19	498	24	2	30	86	583	25

(continued)

Table 10.
Grade 5 RS-to-SS Tables : Op. 1 (CR2), Op. 2 (Weighted CR1), and Op. 3 (Unweighted CR1)

		Option 1 : CR2		Op. 2 : Weighted CR1		Diff. SS	Op. 3 : Unweighted CR1			
RS	% NC	SS	SE	SS	SE	(Op1-Op2)	RS	% NC	SS	SE
31	65	507	19	505	24	2	31	89	598	28
32	67	514	19	512	24	2	32	91	617	31
33	69	521	19	519	23	2	33	94	642	38
34	71	528	19	526	23	2	34	97	682	52
35	73	535	19	534	24	1	35	100	800	163
36	75	543	20	541	24	2				
37	77	551	20	549	25	2				
38	79	560	21	558	25	2				
39	81	570	22	567	26	3				
40	83	580	23	576	27	4				
41	85	592	24	587	29	5				
42	88	605	26	598	31	7				
43	90	621	29	612	34	9				
44	92	640	32	627	37	13				
45	94	663	36	646	42	17				
46	96	693	41	671	51	22				
47	98	738	54	710	73	28				
48	100	800	90	800	172	0				

Table 11.

Grade 8 RS-to-SS Tables : Op. 1 (CR2), Op. 2 (Weighted CR1), and Op. 3 (Unweighted CR1)

	Option 1 : CR2			Op. 2 : Weighted CR1			Op. 3 : Unweighted CR1		
RS	% NC	SS	SE	% NC	SS	SE	% NC	SS	SE
0	0	200	157	0	200	160	0	200	182
1	2	200	157	2	200	160	2	200	182
2	3	200	157	3	200	160	4	200	182
3	5	200	157	5	200	160	7	200	182
4	6	200	157	6	200	160	9	200	182
5	8	200	157	8	200	160	11	200	182
6	10	238	120	9	229	130	13	266	116
7	11	283	75	11	274	86	16	310	72
8	13	307	51	13	300	63	18	334	49
9	15	324	38	14	318	50	20	351	37
10	16	336	31	16	331	42	22	363	31
11	18	346	26	17	342	36	24	373	27
12	19	354	23	19	351	32	27	382	25
13	21	360	21	20	359	28	29	390	25
14	23	367	20	22	365	26	31	399	24
15	24	372	19	23	371	25	33	407	24
16	26	378	19	25	377	24	36	414	23
17	27	383	18	27	382	23	38	422	23
18	29	388	18	28	387	23	40	430	22
19	31	393	18	30	392	23	42	437	21
20	32	398	18	31	397	23	44	444	20
21	34	404	19	33	402	23	47	450	20
22	35	409	19	34	407	23	49	457	19
23	37	414	19	36	413	23	51	463	19
24	39	420	19	38	418	23	53	470	19
25	40	425	18	39	423	23	56	476	19
26	42	430	18	41	428	23	58	482	18
27	44	435	18	42	433	22	60	489	18
28	45	440	18	44	438	22	62	495	18
29	47	445	17	45	443	22	64	501	18
30	48	450	17	47	448	21	67	508	18
31	50	455	17	48	453	21	69	515	19

(continued)

Table 11.
Grade 8 RS-to-SS Tables : Op. 1 (CR2), Op. 2 (Weighted CR1), and Op. 3 (Unweighted CR1)

	Option 1 : CR2			Op. 2 : Weighted CR1			Op. 3 : Unweighted CR1		
RS	% NC	SS	SE	% NC	SS	SE	% NC	SS	SE
32	52	460	17	50	458	21	71	522	19
33	53	465	17	52	463	21	73	529	19
34	55	470	17	53	468	21	76	536	20
35	56	475	17	55	472	20	78	544	20
36	58	480	17	56	477	20	80	553	21
37	60	485	17	58	482	20	82	562	22
38	61	490	17	59	487	20	84	573	24
39	63	495	17	61	491	20	87	585	26
40	65	500	17	63	496	20	89	599	29
41	66	505	17	64	501	20	91	616	32
42	68	510	17	66	506	20	93	638	36
43	69	516	17	67	511	21	96	665	42
44	71	521	17	69	516	21	98	708	57
45	73	527	17	70	521	21	100	800	133
46	74	533	18	72	527	21			
47	76	539	18	73	532	22			
48	77	545	18	75	538	22			
49	79	552	19	77	544	23			
50	81	559	20	78	550	23			
51	82	567	21	80	556	24			
52	84	575	22	81	563	25			
53	85	584	23	83	571	26			
54	87	595	25	84	579	28			
55	89	607	27	86	588	30			
56	90	620	28	88	598	31			
57	92	636	31	89	609	33			
58	94	654	33	91	621	36			
59	95	675	36	92	635	38			
60	97	703	42	94	652	41			
61	98	745	59	95	671	46			
62	100	800	98	97	697	54			
63				98	737	77			
64				100	800	140			

Table 12.
Scale-Score Comparisons of the Three Options : ER Context

Grade 3 : ER forms

Scale scores:	Mean	SD		
Option 1 : ER2	497.5	66.6		
Option 2 : Weighted ER1	497.4	68.6		
Option 3 : Unweighted ER1	497.3	69.3		
Difference scores:	Mean	SD	% students w/ diff. ≤ |5|	% students w/ diff. ≤ |10|
Option 2 SS - Option 1 SS	-0.06	13.10	66%	82%
Option 3 SS - Option 1 SS	-0.20	12.07	61%	86%

Grade 5 : ER forms

Scale scores:	Mean	SD		
Option 1 : ER2	499.7	59.5		
Option 2 : Weighted ER1	499.4	58.7		
Option 3 : Unweighted ER1	499.9	61.0		
Difference scores:	Mean	SD	% students w/ diff. ≤ |5|	% students w/ diff. ≤ |10|
Option 2 SS - Option 1 SS	-0.30	9.26	66%	84%
Option 3 SS - Option 1 SS	0.16	8.01	80%	89%

Grade 8 : ER forms

Scale scores:	Mean	SD		
Option 1 : ER2	488.2	68.6		
Option 2 : Weighted ER1	488.6	69.1		
Option 3 : Unweighted ER1	488.7	70.4		
Difference scores:	Mean	SD	% students w/ diff. ≤ |5|	% students w/ diff. ≤ |10|
Option 2 SS - Option 1 SS	0.43	10.32	60%	86%
Option 3 SS - Option 1 SS	0.51	10.11	75%	88%

Table 13.
Scale-Score Comparisons of the Three Options : CR Context

Grade 3 : CR forms

Scale scores:	Mean	SD		
Option 1 : CR2	501.2	64.4		
Option 2 : Weighted CR1	500.9	68.2		
Option 3 : Unweighted CR1	501.0	68.5		
Difference scores:	Mean	SD	% students w/ diff. ≤ |5|	% students w/ diff. ≤ |10|
Option 2 SS - Option 1 SS	-0.28	17.94	35%	63%
Option 3 SS - Option 1 SS	-0.25	17.16	42%	64%

Grade 5 : CR forms

Scale scores:	Mean	SD		
Option 1 : CR2	499.5	59.7		
Option 2 : Weighted CR1	499.7	62.7		
Option 3 : Unweighted CR1	500.0	62.1		
Difference scores:	Mean	SD	% students w/ diff. ≤ |5|	% students w/ diff. ≤ |10|
Option 2 SS - Option 1 SS	0.25	17.13	30%	53%
Option 3 SS - Option 1 SS	0.54	15.71	33%	59%

Grade 8 : CR forms

Scale scores:	Mean	SD		
Option 1 : CR2	488.7	65.3		
Option 2 : Weighted CR1	489.3	67.9		
Option 3 : Unweighted CR1	489.9	67.6		
Difference scores:	Mean	SD	% students w/ diff. ≤ |5|	% students w/ diff. ≤ |10|
Option 2 SS - Option 1 SS	0.54	14.61	34%	54%
Option 3 SS - Option 1 SS	1.12	13.69	37%	63%

Figure 1.
Grade 3 TCCs : Op. 1 (ER2), Op. 2 (Weighted ER1), & Op. 3 (Unweighted ER1)

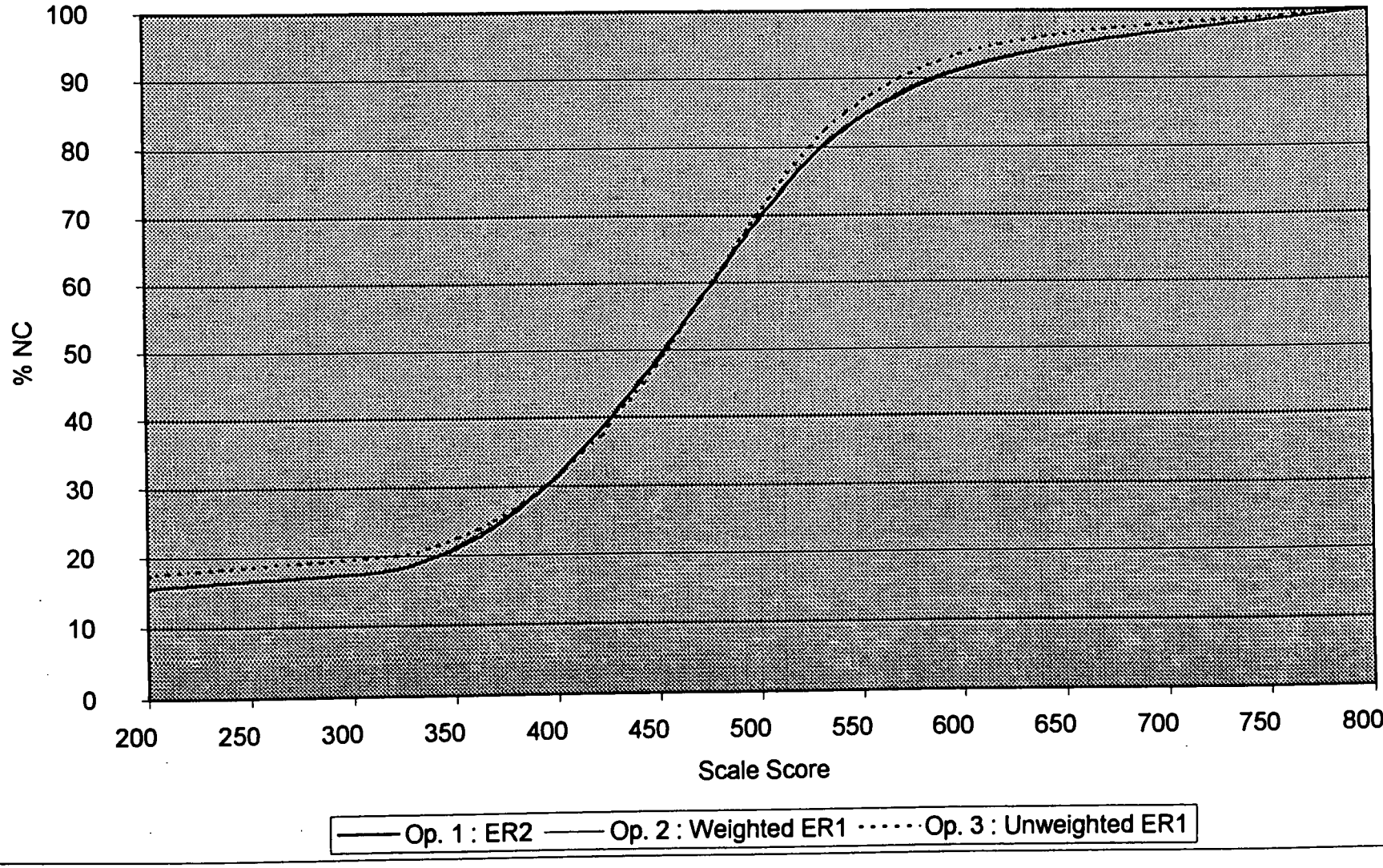


Figure 2.
Grade 5 TCCs: Op. 1 (ER2), Op. 2 (Weighted ER1), & Op. 3 (Unweighted ER1)

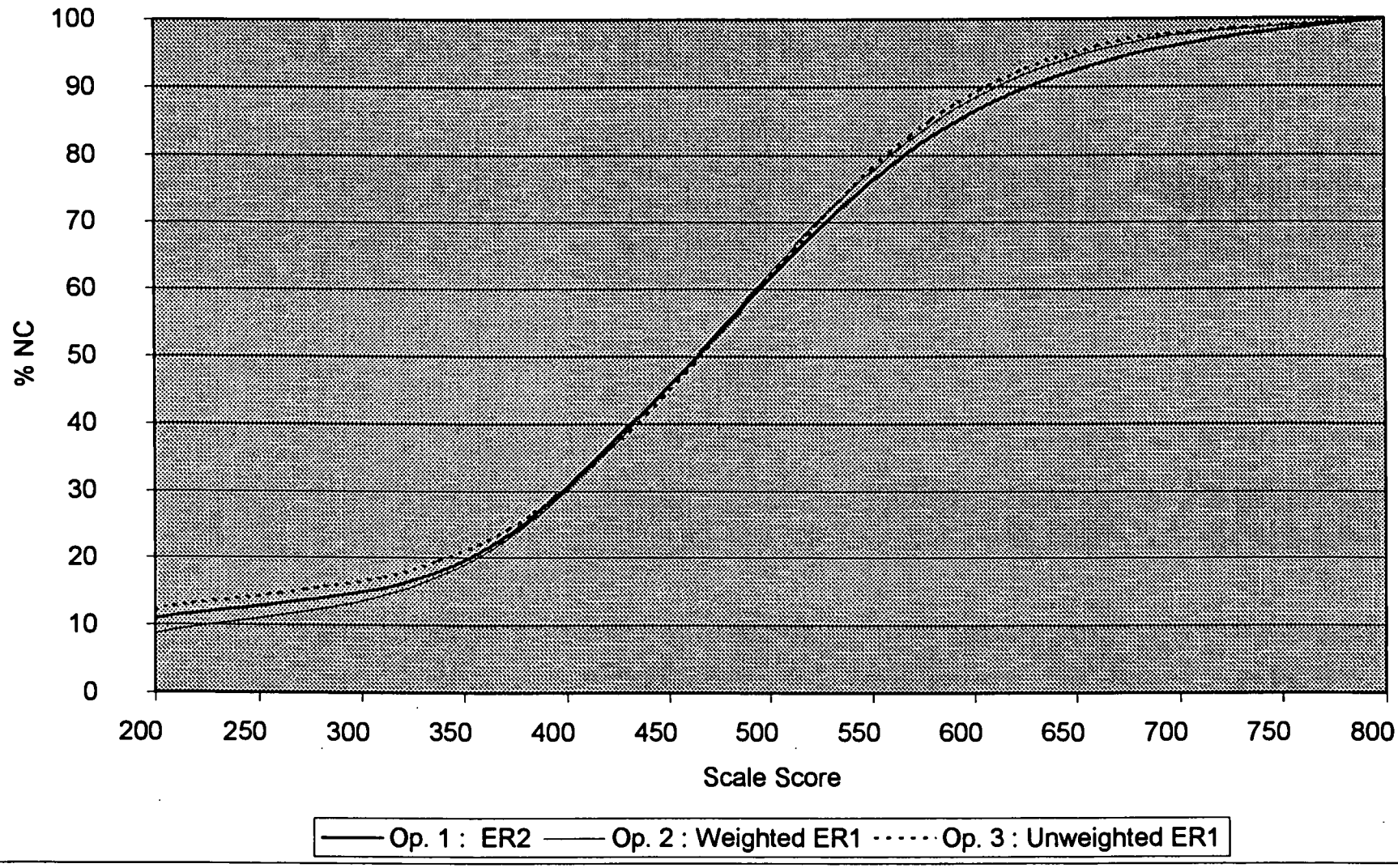
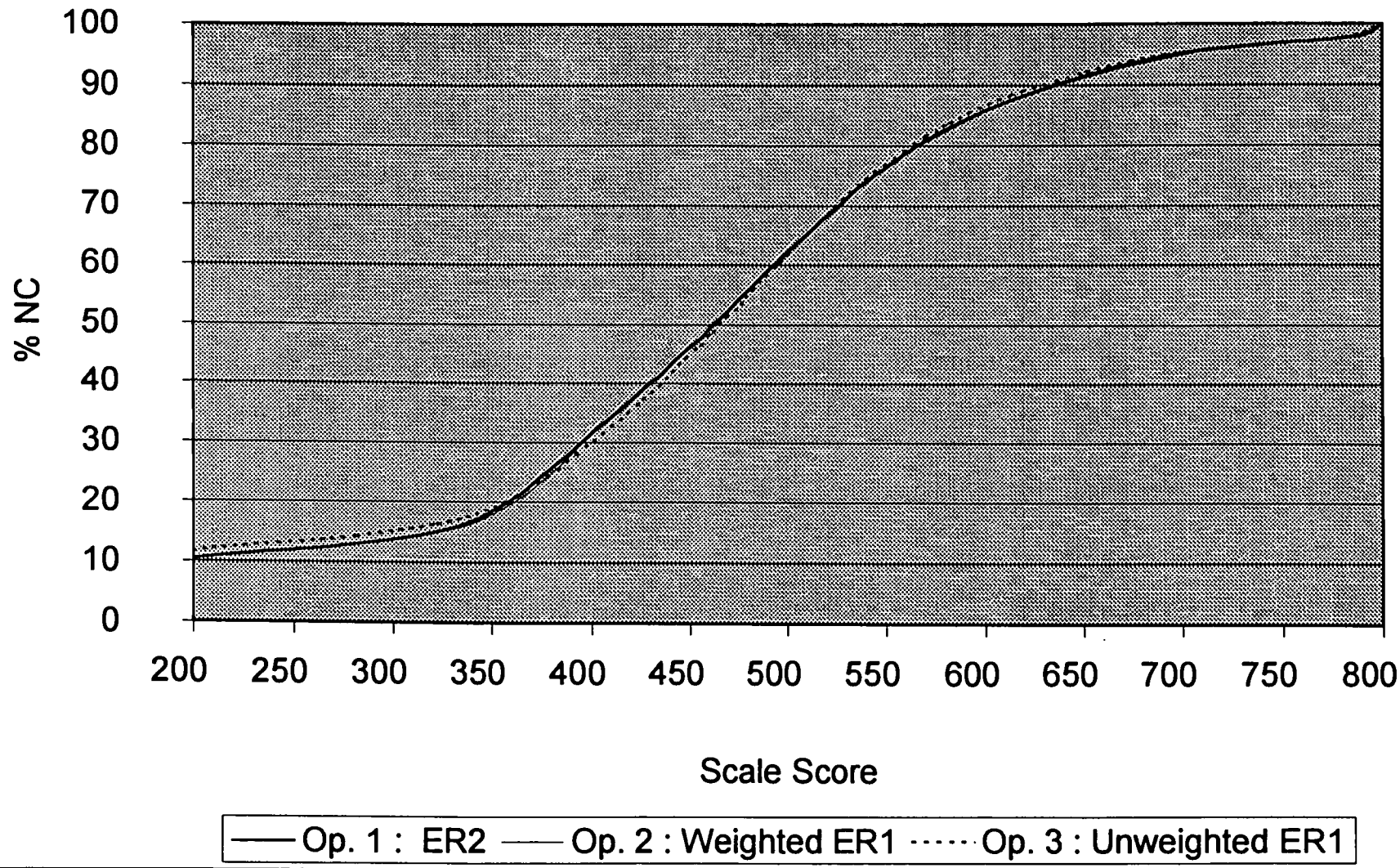


Figure 3.
Grade 8 TCCs : Op. 1 (ER2), Op. 2 (Weighted ER1), & Op. 3 (Unweighted ER1)



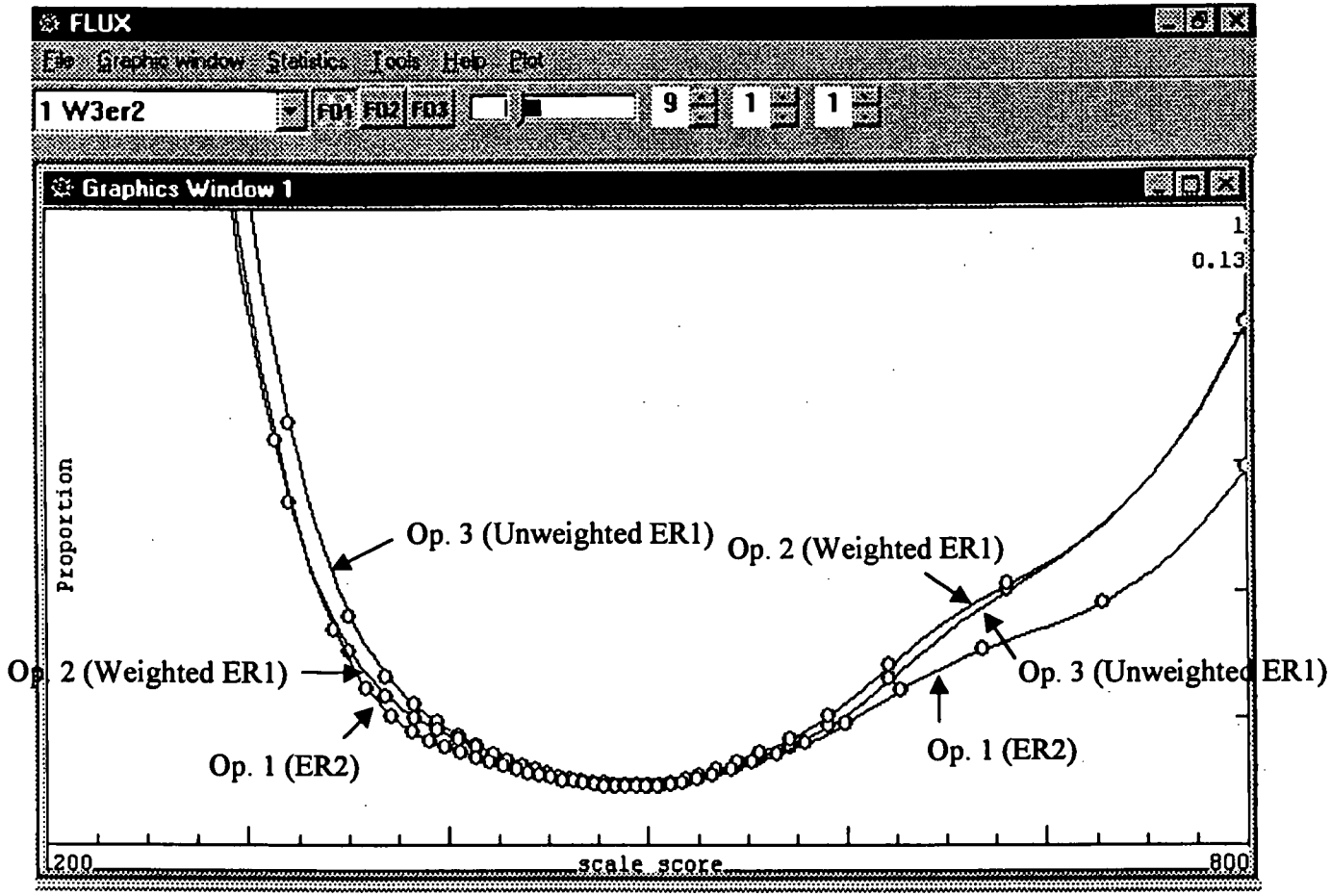


Figure 4.

Grade 3 SE Curves : Op. 1 (ER2), Op. 2 (Weighted ER1), and Op.3 (Unweighted ER1)

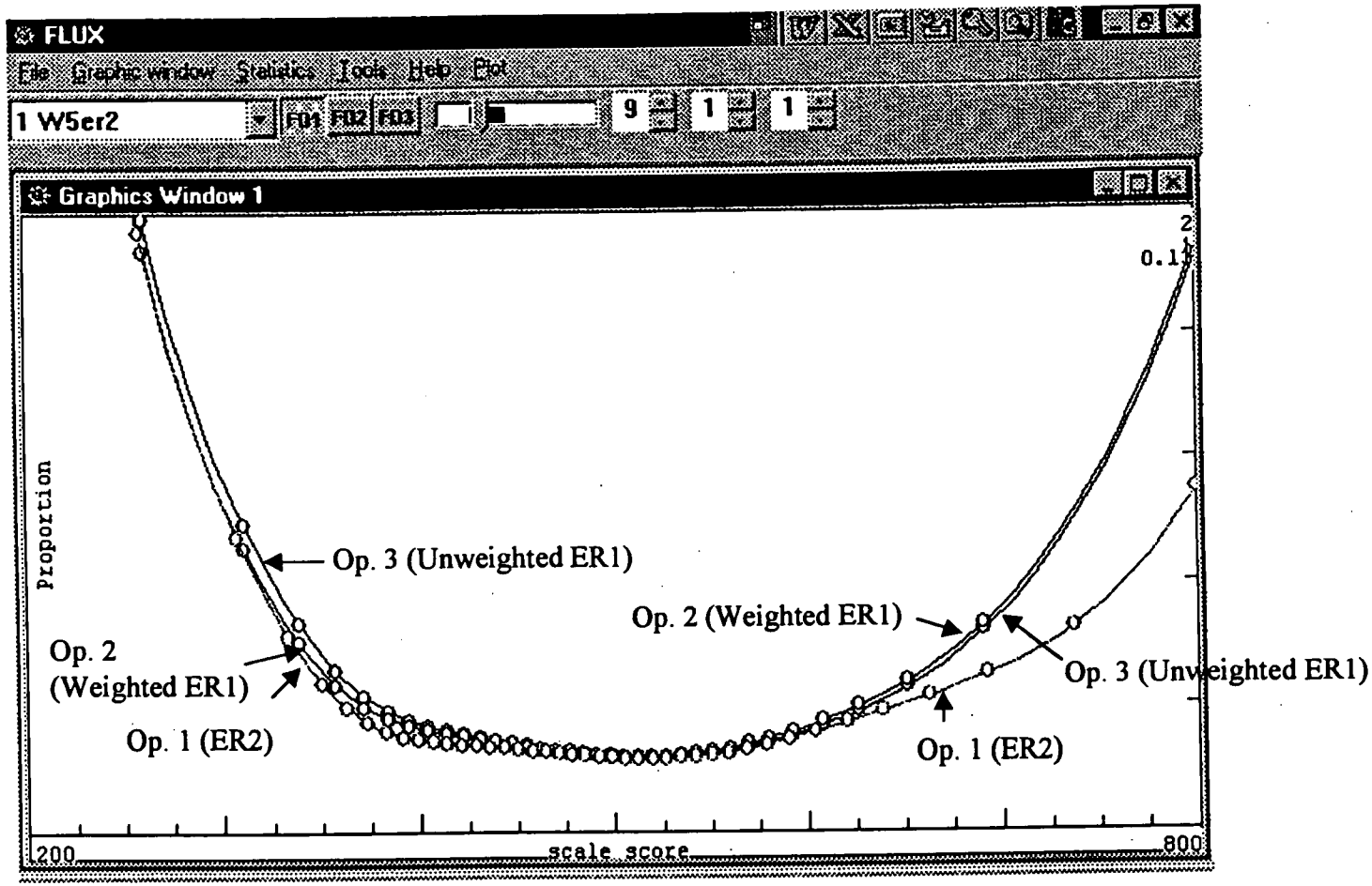


Figure 5.

Grade 5 SE Curves : Op.1 (ER2), Op. 2 (Weighted ER1), and Op. 3 (Unweighted ER1)

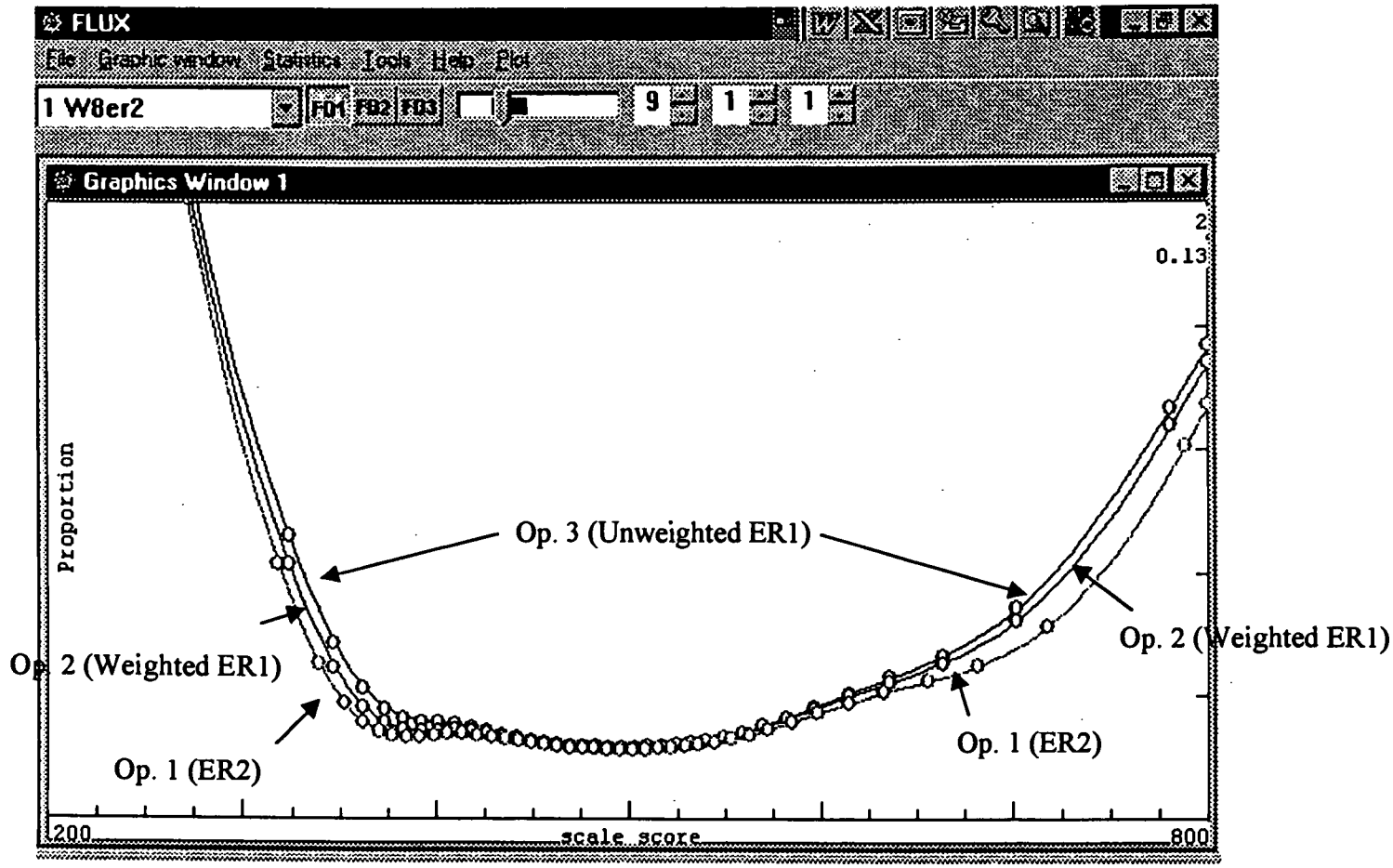


Figure 6.

Grade 8 SE Curves : Op. 1 (ER2), Op. 2 (Weighted ER1), and Op. 3 (Unweighted ER1)

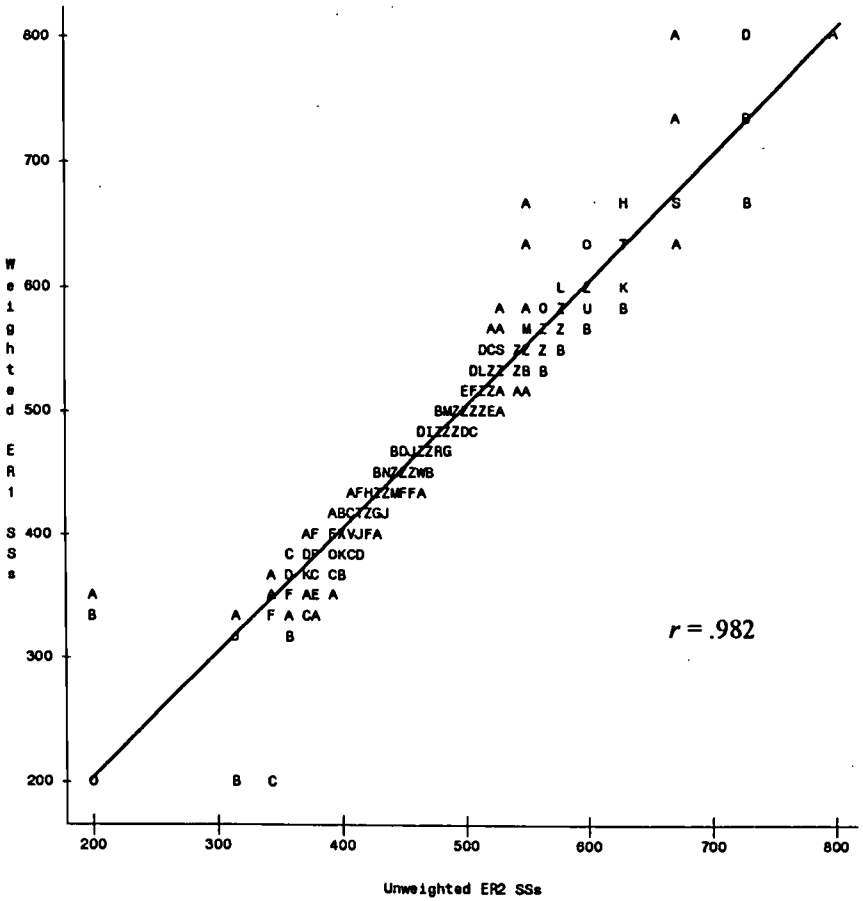


Figure 7.

Grade 3 : Scale-score plot of Op. 1 ER2) SSs and Op. 2 (weighted ER1) SSs

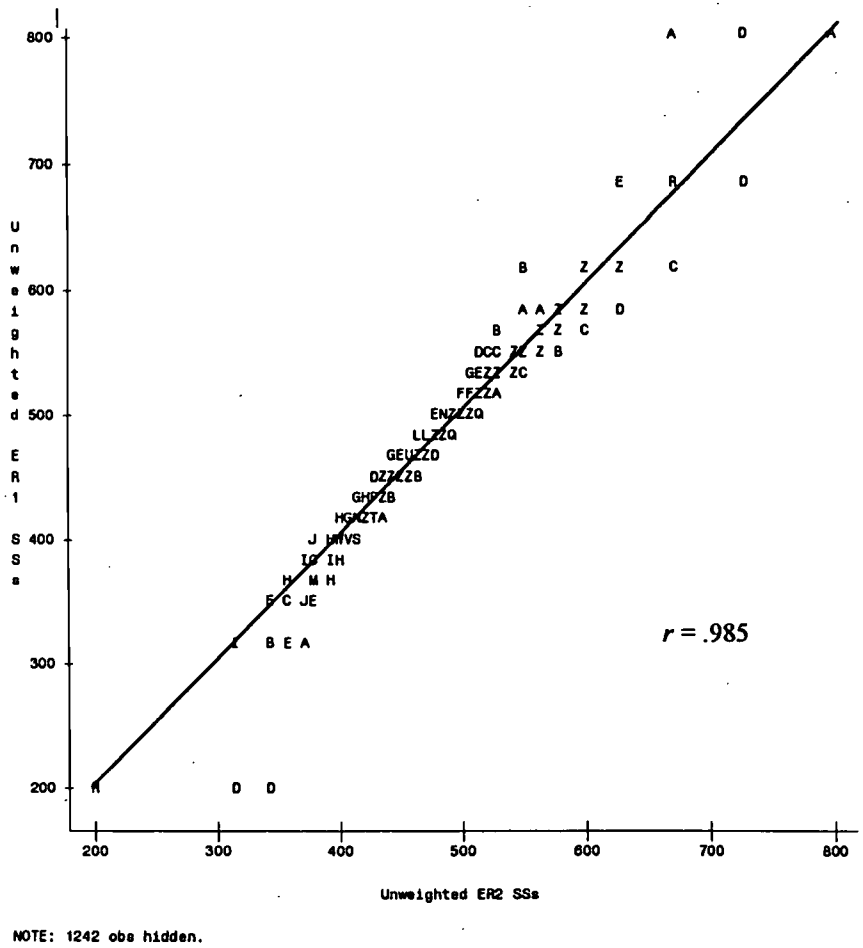
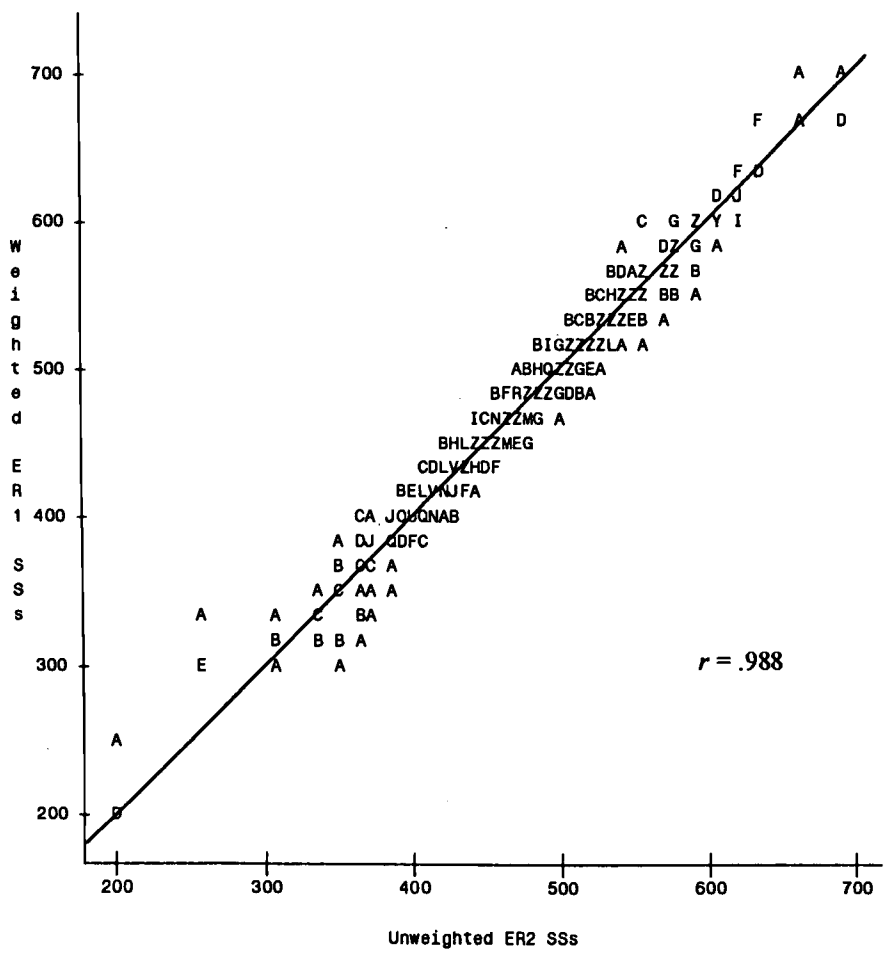


Figure 8.

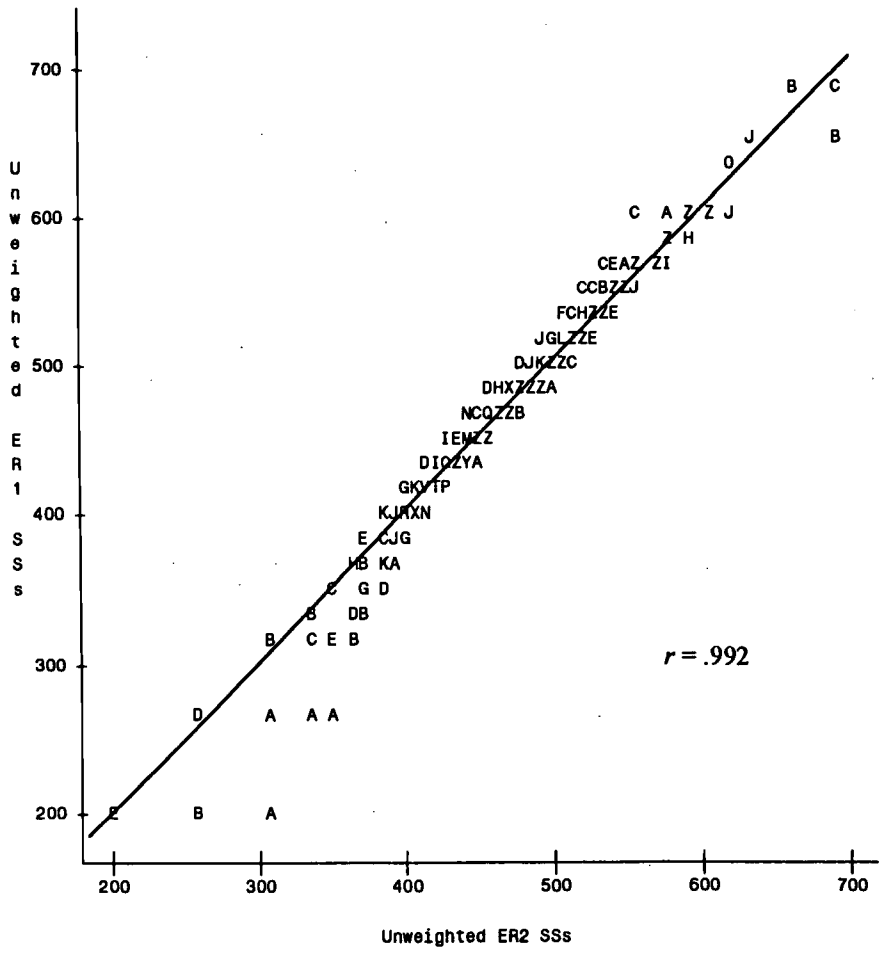
Grade 3 : Scale-score plot of Op.1 (ER2) SSs and Op.3 (unweighted ER1) SSs



NOTE: 1014 obs hidden.

Figure 9.

Grade 5 : Scale-score plot of Op. 1 (ER2) SSs and Op.2 (weighted ER1 SSs)



NOTE: 1211 obs hidden.

Figure 10.

Grade 5 : Scale-score plot of Op. 1 (ER2) SSs and Op. 3 (unweighted ER1) SSs

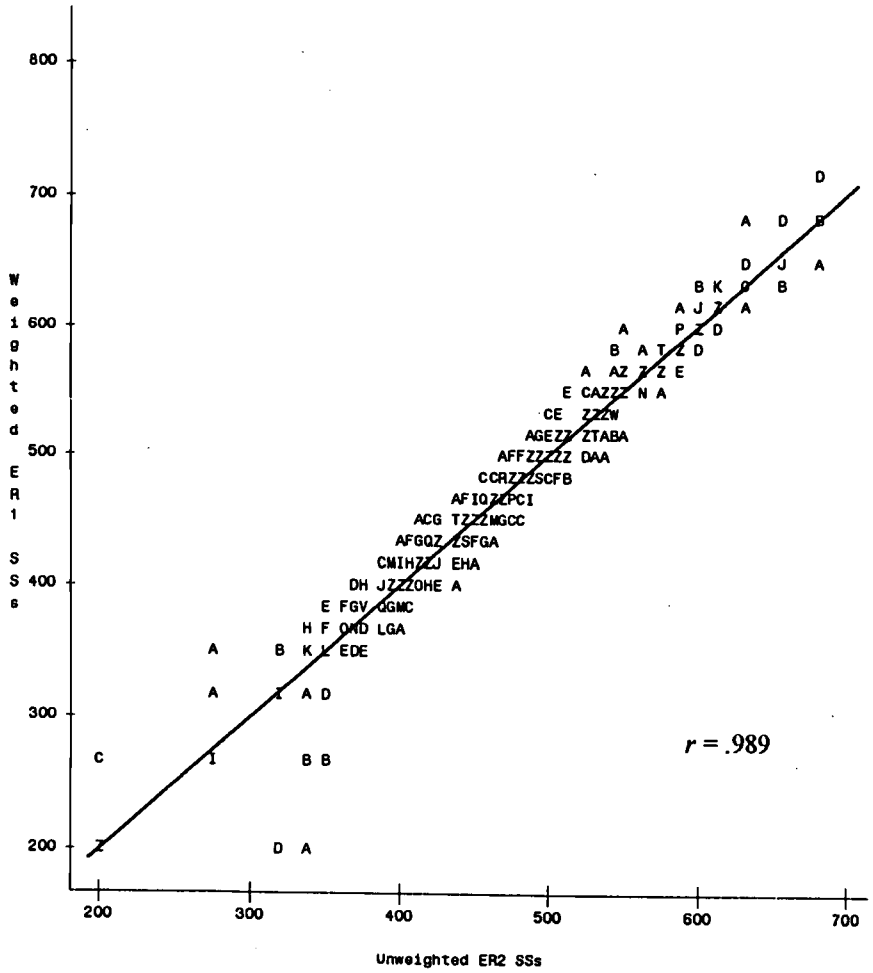


Figure 11.

Grade 8 : Scale-score plot of Op. 1 (ER2) SSs and Op. 2 (weighted ER1) SSs

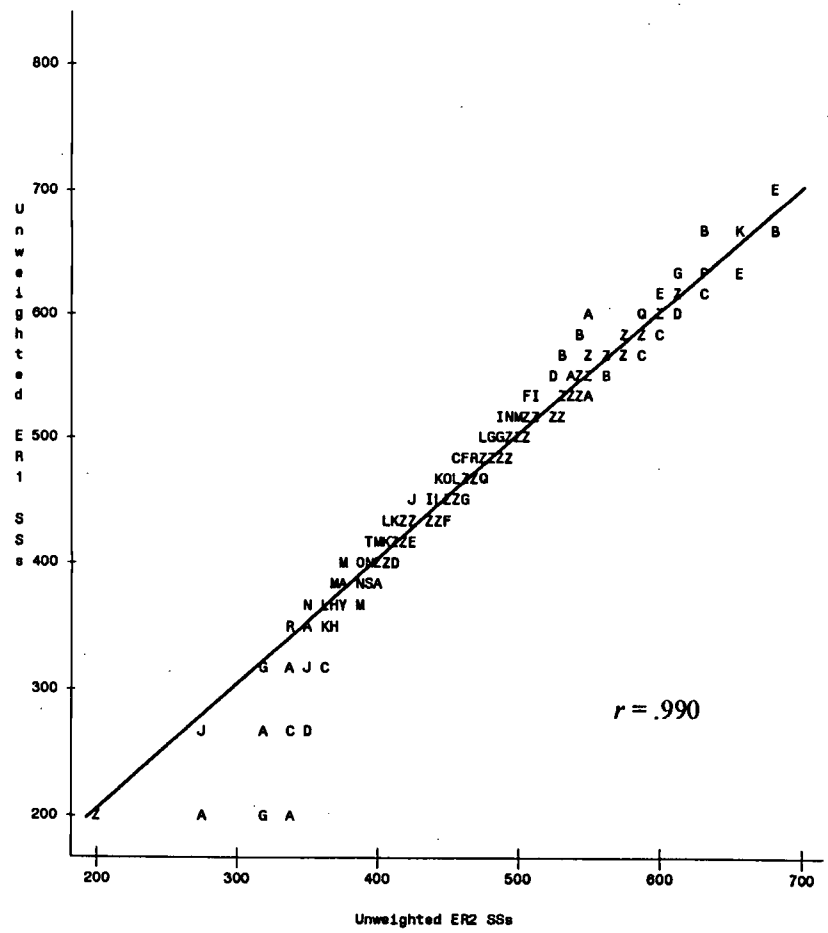


Figure 12.

Grade 8 : Scale-score plot of Op. 1 (ER2) SSs and Op. 3 (unweighted ER1) SSs

Figure 13.
Grade 3 TCCs : Op. 1 (CR2), Op. 2 (Weighted CR1), & Op. 3 (Unweighted CR1)

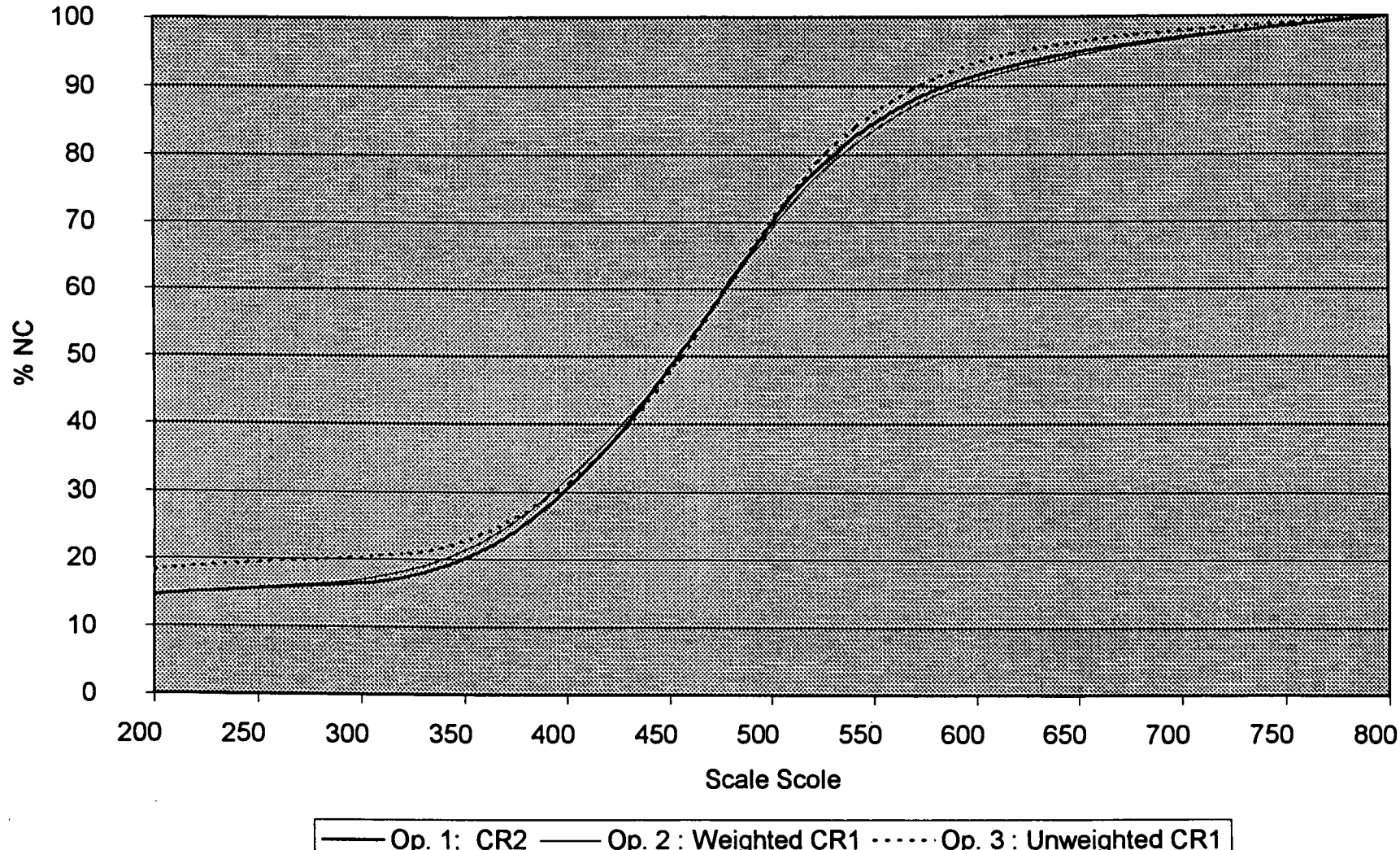


Figure 14.
Grade 5 TCCs : Op. 1 (CR2), Op. 2 (Weighted CR1), & Op. 3 (Unweighted CR1)

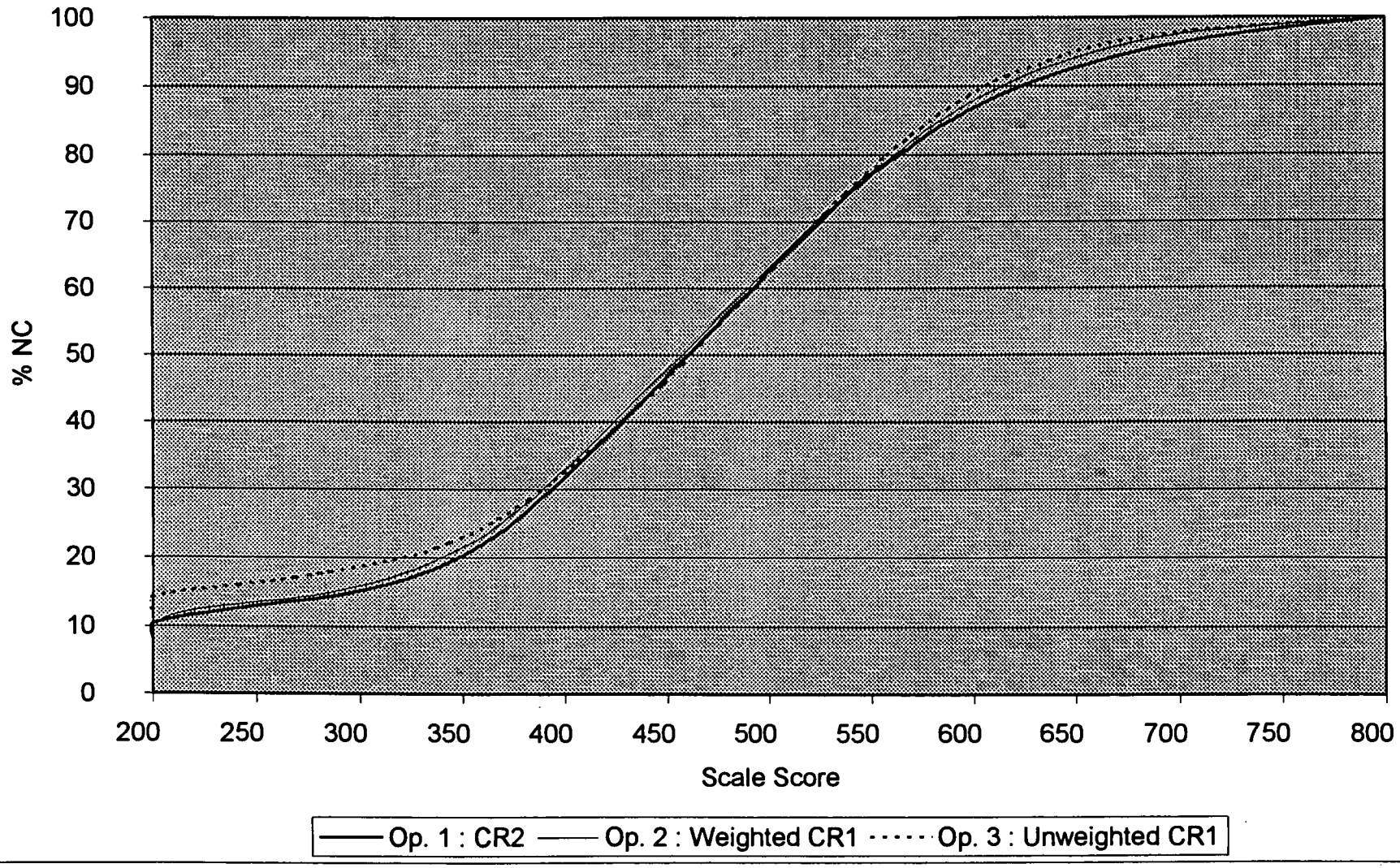
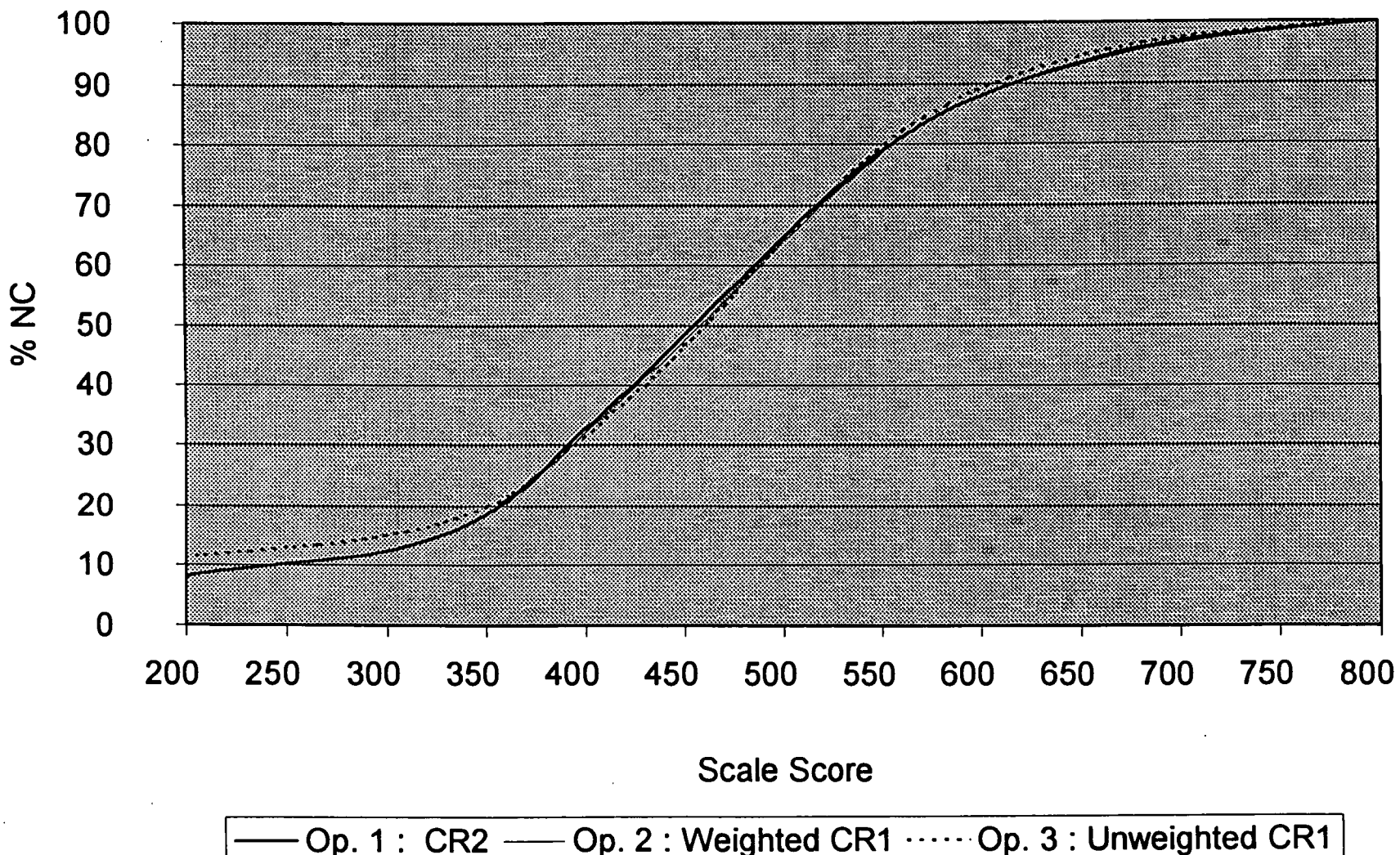


Figure 15.
Grade 8 TCCs : Op. 1 (CR2), Op. 2 (Weighted CR1), & Op. 3 (Unweighted CR1)



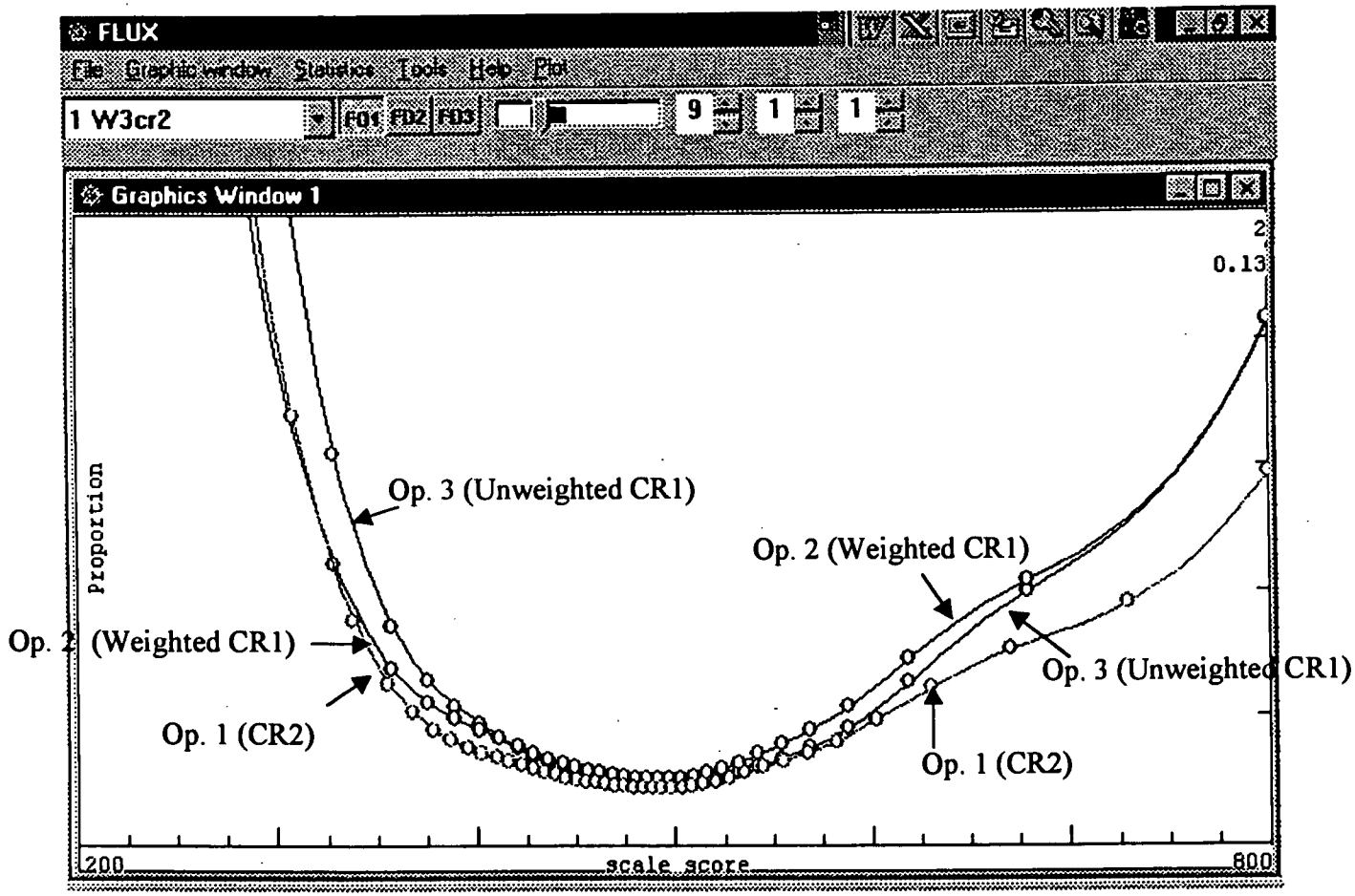


Figure 16.

Grade 3 SE Curves : Op. 1 (CR2), Op. 2 (Weighted CR1), and Op. 3 (Unweighted CR1)

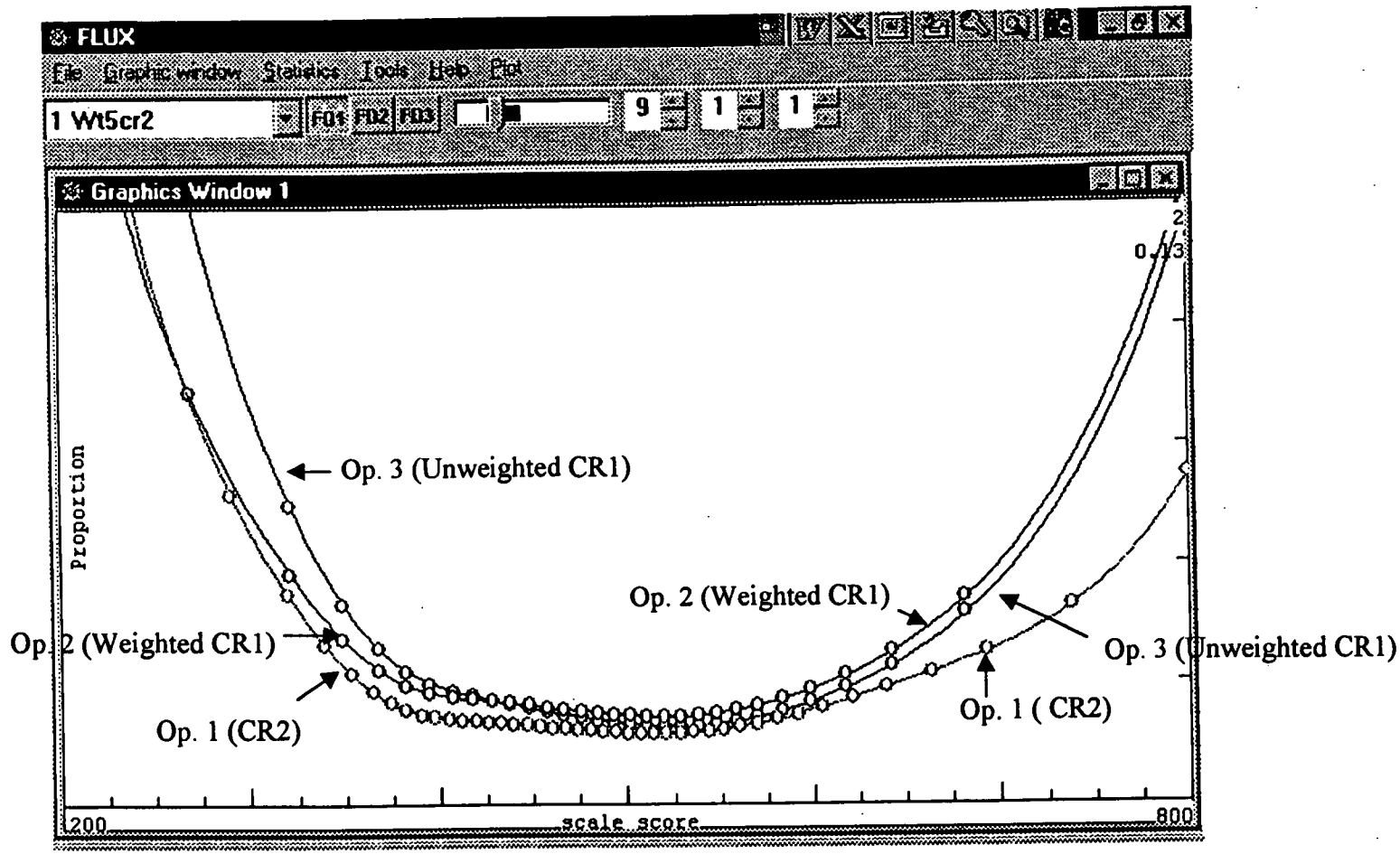


Figure 17.

Grade 5 SE Curves : Op. 1 (CR2), Op. 2 (Weighted CR1), and Op. 3 (Unweighted CR1)

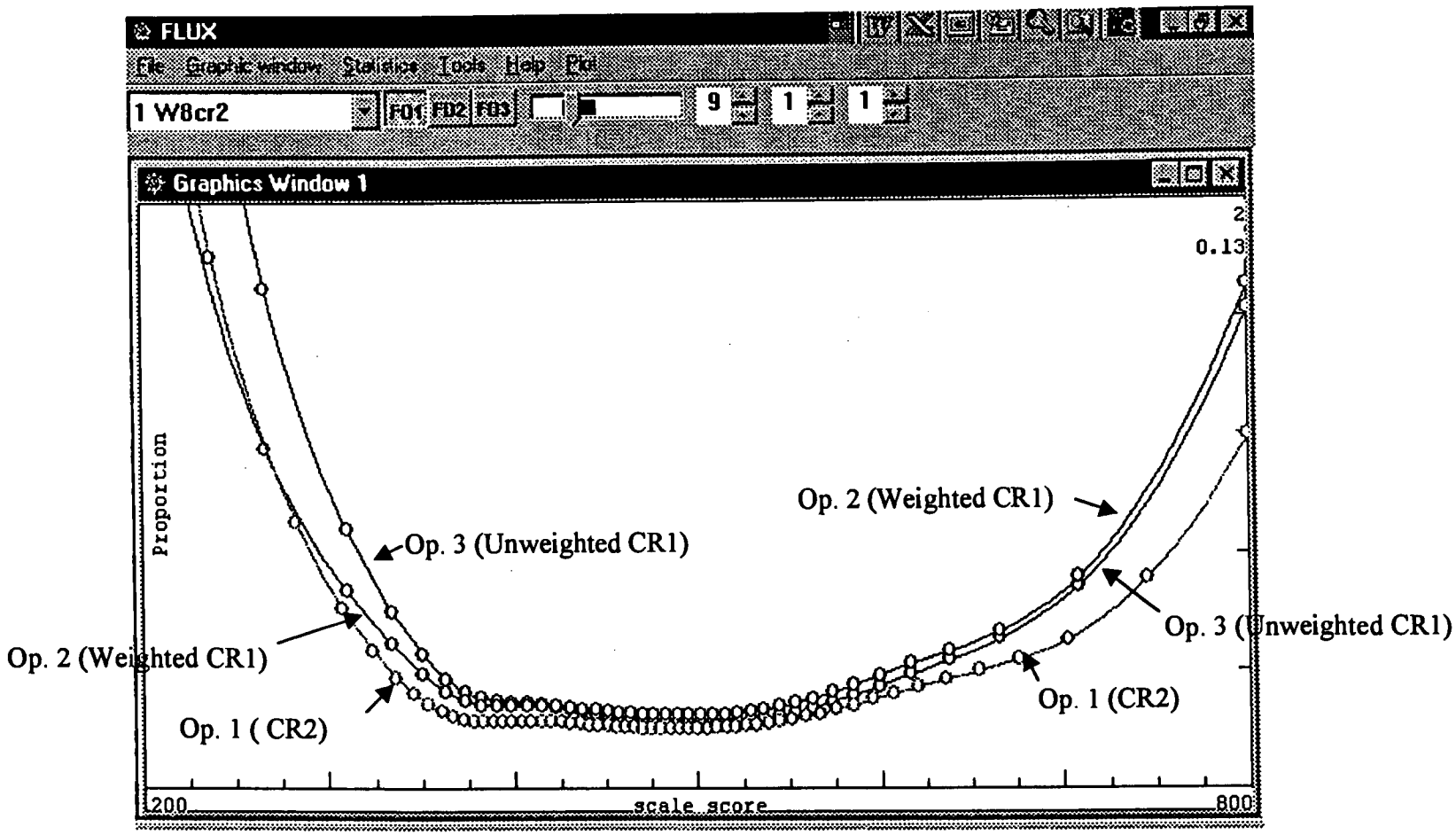


Figure 18.

Grade 8 SE Curves : Op. 1 (CR2), Op. 2 (Weighted CR1), and Op. 3 (Unweighted CR1)

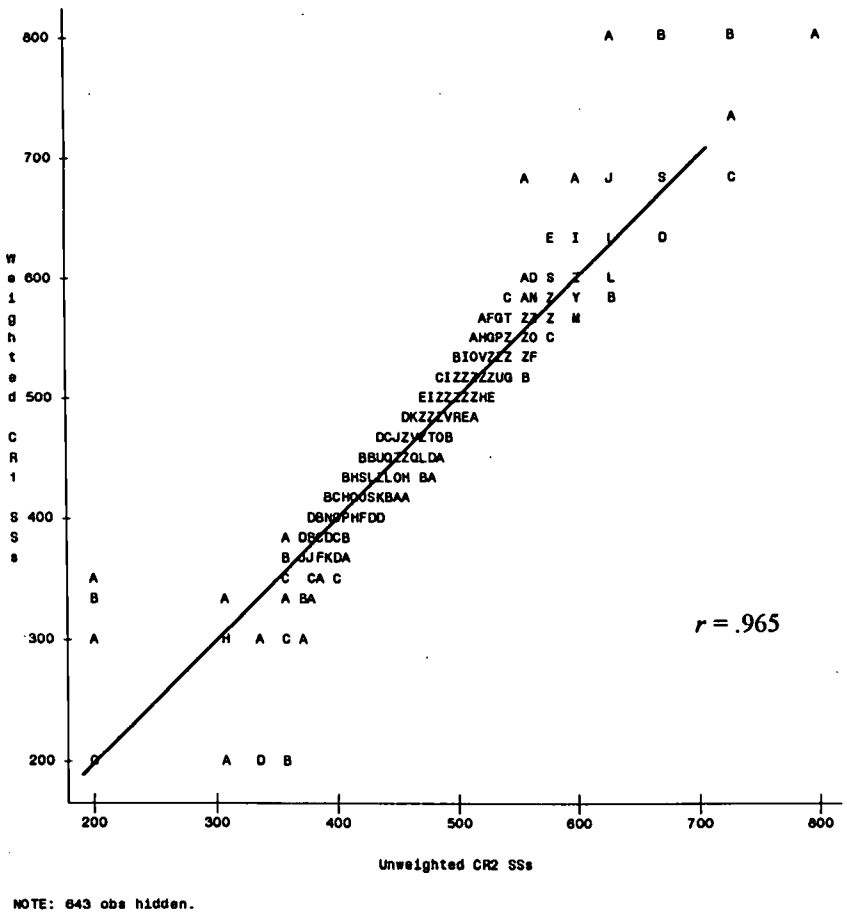


Figure 19.

Grade 3 : Scale-score plot of Op. 1 (CR2) SSs and Op. 2 (weighted CR1) SSs

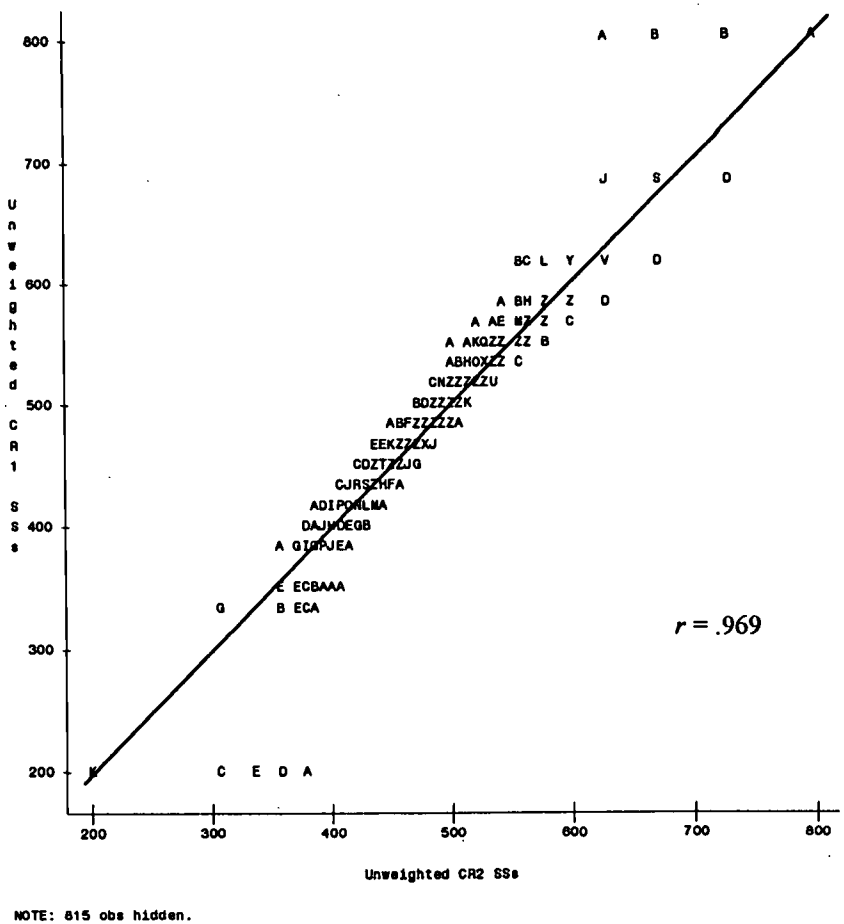
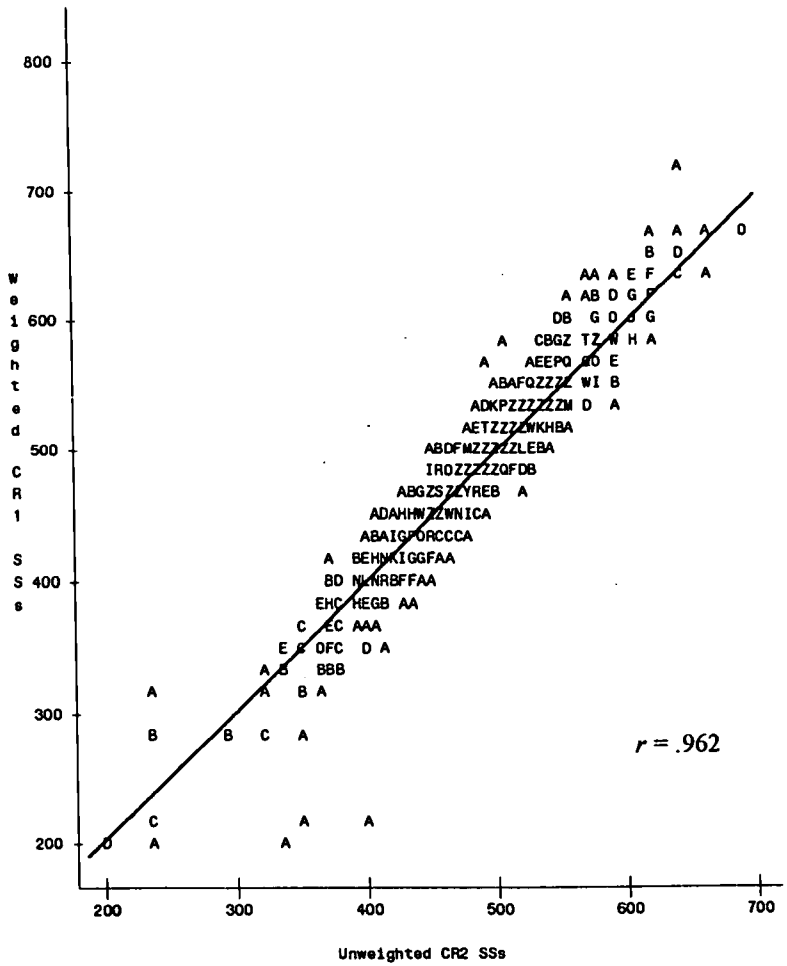


Figure 20.

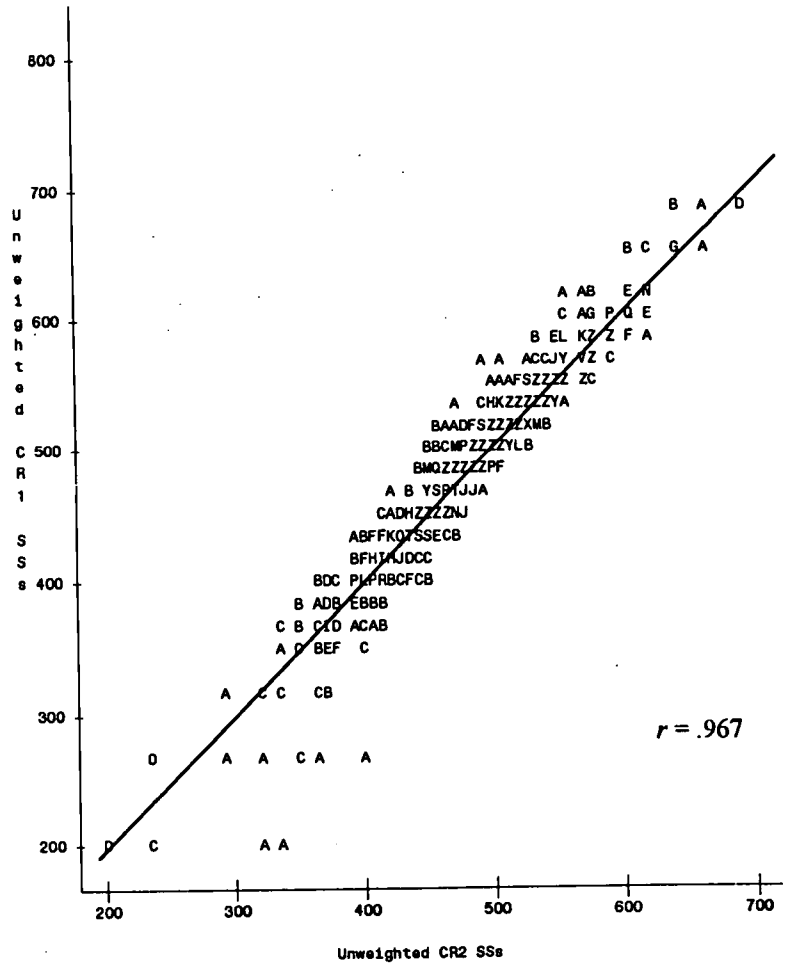
Grade 3 : Scale-score plot of Op. 1 (CR2) SSs and Op. 3 (unweighted CR1) SSs



NOTE: 467 obs hidden.

Figure 21.

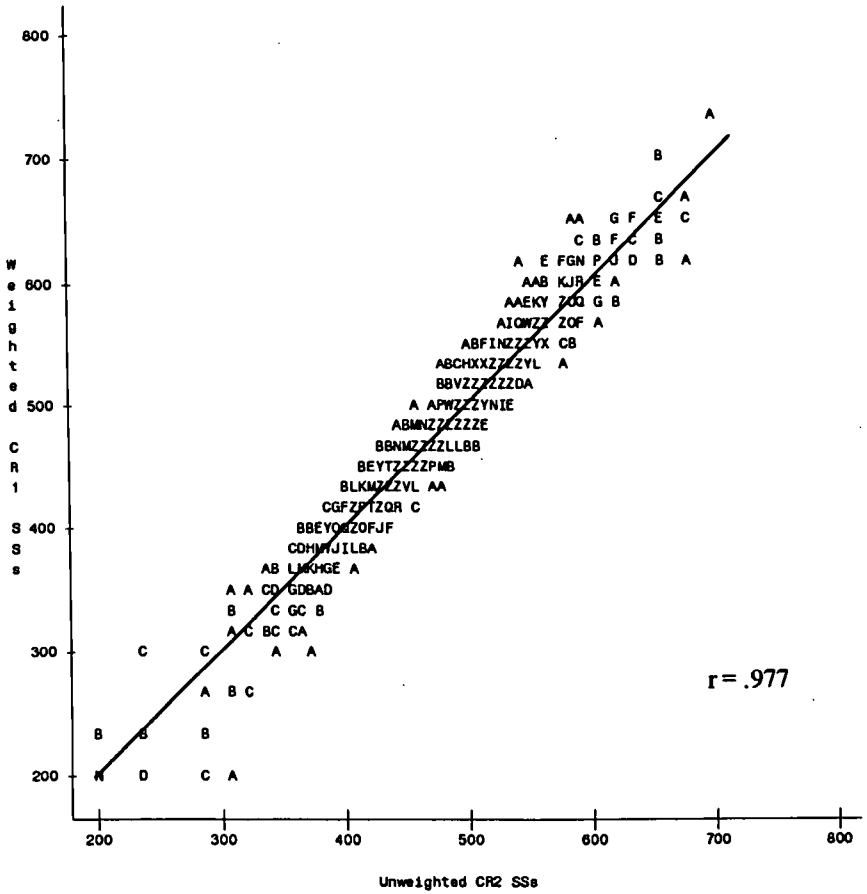
Grade 5 : Scale-score plot of Op. 1 (CR2) SSs and Op. 2 (weighted CR1) SSs



NOTE: 584 obs hidden.

Figure 22.

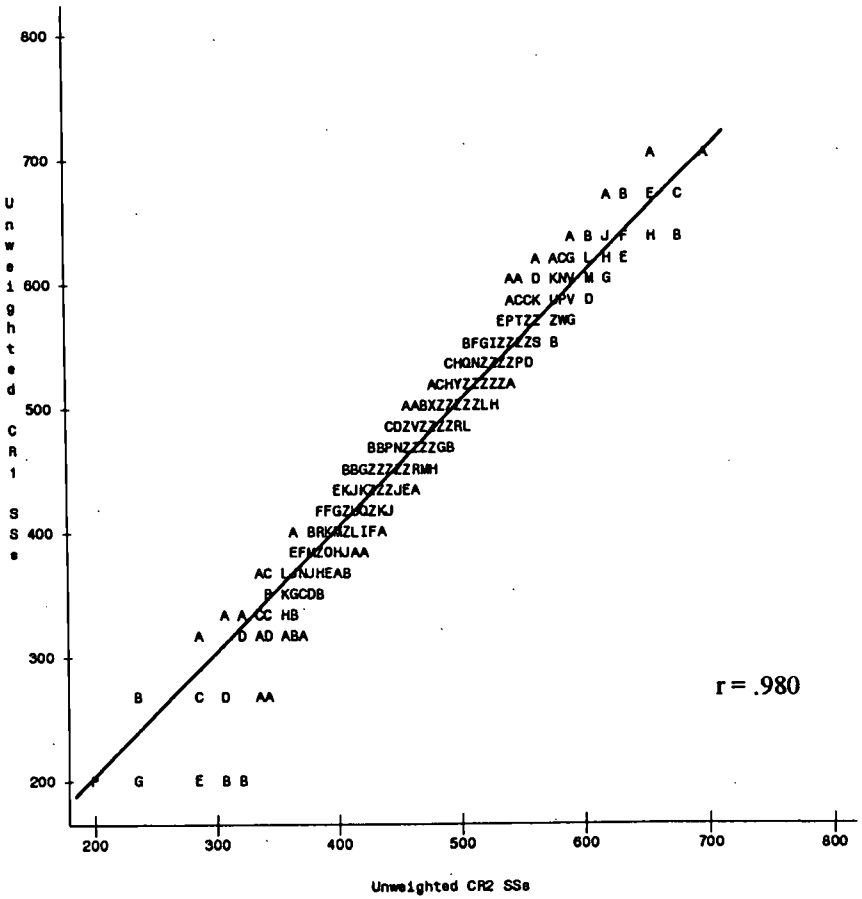
Grade 5 : Scale-score plot of Op. 1 (CR2) SSs and Op. 3 (unweighted CR1) SSs



NOTE: 1119 obs hidden.

Figure 23.

Grade 8 : Scale-score plot of Op. 1 (CR2) SSs and Op. 2 (weighted CR1) SSs



NOTE: 1309 obs hidden.

Figure 24.

Grade 8 : Scale-score plot of Op. 1 (CR2) SSs and Op. 3 (unweighted CR1) SSs

MISSISSIPPI GRADE LEVEL TESTING PROGRAM
CRT FALL 2000 LANGUAGE BLUE PRINT'S

Language	Total Number of Items						
Reporting Categories	**Grade 2**	**Grade 3**	**Grade 4**	**Grade 5**	**Grade 6**	**Grade 7**	**Grade 8**
Editing: Capitalization and Punctuation	19 MC 2 OE	14 MC 2 OE	13 MC 2 OE	13 MC 2 OE	13 MC 2 OE	13 MC 2 OE	13 MC 2 OE
Spelling	4 MC	4 MC	10 MC	10 MC	10 MC	10 MC	10 MC
Language Structure (Syntactic)	14 MC 1 OE	16 MC 1 OE	14 MC 1 OE	15 MC	15 MC	16 MC	15 MC
Meaning (Semantic)	12 MC 2 OE	16 MC 2 OE	13 MC 2 OE	12 MC 3 OE	12 MC 3 OE	11 MC 3 OE	14 MC 3 OE

MISSISSIPPI GRADE LEVEL TESTING PROGRAM
CRT FALL 2000 LANGUAGE BLUE PRINT'S

Language	Benchmarks/Items						
Reporting Categories	**Grade 2**	**Grade 3**	**Grade 4**	**Grade 5**	**Grade 6**	**Grade 7**	**Grade 8**
Editing: Capitalization and Punctuation	5 6 16 17	4 16	2 4	2 4	4 5 23 25	4 6 24 25	4 5
Spelling	11 12 13 16 17	9 16	2 4	4	5 23 25	5 6	5
Language Structure (Syntactic)	4 16 17	3 4 6 8	1 3	1 3 4	1 2 3 5	1 2 3 6	1 2 4 7
Meaning (Semantic)	5 6 7 8 10 16 17 20	5 6 15 18 20 21	5 6 7 8 15 16 17 18	5 6 7 8 9 10 11 14 15 16 17 18 19 20	3 6 9 10 18 19 20 22 23 24 25 26 27 28	2 6 7 9 10 11 12 13 14 15 17 18 19 20 21 22 23 24 25	2 5 6 7 9 10 11 12 13 14 16 17 19 20 22 23

MISSISSIPPI GRADE LEVEL TESTING PROGRAM
CRT FALL 2000 MATHEMATICS BLUE PRINT'S

Mathematics	Total Number of Items						
Reporting Categories	**Grade 2**	**Grade 3**	**Grade 4**	**Grade 5**	**Grade 6**	**Grade 7**	**Grade 8**
Patterns, Algebraic Thinking	7 MC 1 OE	7 MC 1 OE	5 MC 1 OE	6 MC	7 MC 1 OE	9 MC 1 OE	9 MC 1 OE
Data Analysis Prediction	7 MC 1 OE	7 MC 1 OE	7 MC 1 OE	7 MC 1 OE	7 MC 1 OE	7 MC 1 OE	7 MC 1 OE
Measurement	9 MC 1 OE	9 MC 1 OE	8 MC 1 OE	8 MC 1 OE	8 MC 1 OE	5 MC 1 OE	5 MC 1 OE
Geometric Concepts	6 MC 1 OE	6 MC	7 MC 1 OE	9 MC 1 OE	8 MC 1 OE	9 MC 1 OE	9 MC 1 OE
Number Sense	21 MC 1 OE	21 MC 2 OE	23 MC 1 OE	20 MC 2 OE	20 MC 1 OE	20 MC 1 OE	20 MC 1 OE

MISSISSIPPI GRADE LEVEL TESTING PROGRAM
CRT FALL 2000 MATHEMATICS BLUE PRINT'S

Mathematics	Benchmarks/Items						
Reporting Categories	**Grade 2**	**Grade 3**	**Grade 4**	**Grade 5**	**Grade 6**	**Grade 7**	**Grade 8**
Patterns, Algebraic Thinking	5b 1b 6b 6j	1a 1b 1c 8a 8b 8c	1a 1b 1c	5j	1a 1b 1c 1d 1e 1f	7a 7b 7c 7d 7e	3a 3b 3c 3d 3e 3f 3g 3h 8a 8b 8c 8d 3a 3b 3c 3d 3e 3f 3g 3h
Data Analysis Predication	3a 3b 3c 3d	4a 4b 4c 4d 4e	4a 4b 4c	3a 3b 3c 3d	5a 5b 5c 5d 5e 5f	4a 4b 4c 4d 4e 4f 4g 4h	7a 7b 7c 7d 7e 7f 7g

MISSISSIPPI GRADE LEVEL TESTING PROGRAM
CRT FALL 2000 MATHEMATICS BLUE PRINT'S

Measurement	2a 2b 2c 2d 2e 4a 4b 4c 4d 4e 5a 5b 5c 5d 5e 5f 5g	3a 3b 3c 3d 3e 3f 3g 5a 5b 5c 5d 5e	3a 3b 3c 3d 3e	2a 2b 2c 2d 2e 2f 2g	4a 4b 4c 4d 4e 4f	3a 3b 3c	5a 5b 5c 5d
Mathematics	**Benchmarks/Items**						
Reporting Categories	**Grade 2**	**Grade 3**	**Grade 4**	**Grade 5**	**Grade 6**	**Grade 7**	**Grade 8**
Geometric Concepts	1a 1b 1c 1d 1e	2a 2b 2c 2d 2e	2a 2b 2c 2d 2e 2f 2g 2h	1a 1b 1c 1d 1e 1f 1g	2a 2b 2c 2d 2e 3a 3b 3c 3d	5a 5b 5c 5d 5e 5f 5g 5h 5i 5j 5k 5l	6a 6b 6c 6d 6e 6f 6g 6h 6i

MISSISSIPPI GRADE LEVEL TESTING PROGRAM
CRT FALL 2000 MATHEMATICS BLUE PRINT'S

Number Sense	6a 6b 6c 6d 6e 6f 6g 6h 6i 6j 6k 6l 7a 7b 7c 7d 7e 8a 8b 8c 8d	6a 6b 6c 6d 6e 6f 6g 6h 6i 6j 6k 6l 6m 7a 7b 7c 7d 7e 7f 7g 7h 7i 7j 7k 7l 7m 7n 9a 9b 9c 9d 9e	5a 5b 5c 6a 6b 6c 6d 6e 6f 7a 7b 7c 7d 7e 7f 7g 7h 7i	4a 4b 4c 4d 4e 4f 4g 5a 5b 5c 5d 5e 5f 5g 5h 5i 5j 5k	6a 6b 6c 6d 6e 6f 6g 7a 7b 7c 7d 8a 8b 8c 8d 8e 9a 9b 9c 9d 9e 9f 9g 10a 10b 10c 10d 10e	1a 1b 1c 1d 1e 1f 1g 2a 2b 2c 2d 2e 2f 2g 6a 6b 6c 6d 8a 8b 8c 8d 8e 8f 8g 8h 8i 8j	1a 1b 1c 1d 1e 1f 1g 2a 2b 2c 2d 2e 2f 4a 4b 4c 4d 4e 4f 4g

MISSISSIPPI GRADE LEVEL TESTING PROGRAM
CRT FALL 2000 READING BLUE PRINT'S

Reading	Total Number of Items						
Reporting Categories	**Grade 2**	**Grade 3**	**Grade 4**	**Grade 5**	**Grade 6**	**Grade 7**	**Grade 8**
Context Cues (Semantic)	**5-7 MC 0 CR**	**6 MC 0 CR**	**7-8 MC 0 CR**	**8-9 MC 0 CR**	**6 MC 0 CR**	**5-6 MC 0 CR**	**5-6 MC 0 CR**
Language Structure (Syntactic)	5-7 MC 0-1 CR	6 MC 0 CR	5-8 MC 0 CR	5-6 MC 0 CR	6 MC 0 CR	6 MC 0 CR	6-8 MC 0 CR
Word Patterns (Phonetic Structure)	5-6 MC 0 CR	6 MC 0 CR	Not Assessed	Not Assessed	Not Assessed	Not Assessed	Not Assessed
Vocabulary	7-10 MC 0 CR	6 MC 0 CR	8-9 MC 0 CR	5-6 MC 0 CR	6-7 MC 0 CR	6 MC 0 CR	6 MC 0 CR
Main Idea and Details (Textual)	11-13 MC 2-3 CR	11-15 MC 2-3 CR	8-10 MC 2-3 CR	12-16 MC 2-3 CR	12-15 MC 1-4 CR	9-12 MC 2-3 CR	10-11 MC 2 CR
Extended Meaning/ Thinking (Metacognitive)	9-15 MC 2-3 CR	11-15 MC 2-3 CR	13-15 MC 1-2 CR	10-14MC 2 CR	12-14 MC 10-2 CR	13-15 MC 1-3 CR	10-16 MC 1-2 CR
Workplace Data (Evaluative)	Not Assessed	Not Assessed	4 MC 1-2 CR	4-5 MC 0-1 CR	4-7 MC 0-4 CR	6-8 MC 0-1 CR	5-7 MC 1-2 CR

MISSISSIPPI GRADE LEVEL TESTING PROGRAM
CRT FALL 2000 READING BLUE PRINT'S

Reading	Benchmarks/Items						
Reporting Categories	**Grade 2**	**Grade 3**	**Grade 4**	**Grade 5**	**Grade 6**	**Grade 7**	**Grade 8**
Context Cues (Semantic)	4 8 16 21	4 11 14	7 8 13 15	4 8 11	5 9 14	6	6 7 8 20
Language Structure (Syntactic)	22 13	11 15 10	5 7 9 12	5 8 9	5 7 10	5	5 8
Word Patterns (Phonetic Structure)	7 14 15 6	10 11	5 10 11	5 10 12	5 6 7 11	5 7 8	5 8 3 10
Vocabulary	8 21 22	8 15	7 9 40	6 8 12 15	8 9 11 14	Writing #7	5 6 7 20
Main Idea and Details (Textual)	1 4 18 20 21 24	4 12 14 17 19 20	16 25 17 29 19 19 21 41 22 42 23	16 23 18 25 20 29 21 34 22 36	15 23 14 24 14 30 14 33 22	22 21 22 22 22 23 22 31 20	12 21 14 22 16 23 17 24 18 33 20
Extended Meaning/Thinking (Metacognitive)	4 18 20 23 25 26	4 12 16 18 19	14 22 32 15 28 34 16 29 35 18 30 36 20 31	22 30 38 22 31 22 32 17 33 22 34 22 37	22 27 22 28 22 29 22 30 26 32 34	11 23 13 25 16 26 18 27 22 28	13 24 29 15 26 30 20 27 32 22 28
Workplace Data (Evaluative)	2 3	3 20	16 38 43 17 39 25 40 26 41 37 42	16 42 18 43 27 44 40 41	14 38 14 39 14 40 14 41 37	14 37 15 38 20 39 35 40 36 41	12 37 19 38 21 39 35 36

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

TM031676

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: An evaluation of "intentional" weighting of extended-response or constructed-response items in tests with mixed item types	
Author(s): Kyoko Ito and Robert C. Sykes	
Corporate Source: CTB/McGraw-Hill	Publication Date:

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Level 1

☒

Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.

The sample sticker shown below will be affixed to all Level 2A documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

Level 2A

☐

Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only

The sample sticker shown below will be affixed to all Level 2B documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2B

Level 2B

☐

Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signature: Kyoko Ito	Printed Name/Position/Title: Kyoko Ito, Senior Research Scientist
Organization/Address: CTB/McGraw-Hill, 20 Ryan Ranch Road, Monterey, CA 93940	Telephone: (831) 393-7463 FAX: (831) 393-7616
	E-Mail Address: kito@ctb.com Date: 7/24/00

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

ERIC CLEARINGHOUSE ON ASSESSMENT AND EVALUATION
UNIVERSITY OF MARYLAND
1129 SHRIVER LAB
COLLEGE PARK, MD 20772
ATTN: ACQUISITIONS

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
4483-A Forbes Boulevard
Lanham, Maryland 20706

Telephone: 301-552-4200
Toll Free: 800-799-3742
FAX: 301-552-4700
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com