Listening. Learning. Leading.®

Research Report
ETS RR–11-46

Multiple Linking in Equating and Random Scale Drift

Hongwen Guo

Jinghua Liu

Neil Dorans

Miriam Feigenbaum

December 2011

Multiple Linking in Equating and Random Scale Drift

Hongwen Guo, Jinghua Liu, Neil Dorans, and Miriam Feigenbaum
ETS, Princeton, New Jersey

December 2011

As part of its nonprofit mission, ETS conducts and disseminates the results of research to advance quality and equity in education and assessment for the benefit of ETS's constituents and the field.

To obtain a PDF or a print copy of a report, please visit:

http://www.ets.org/research/contact.html

Technical Review Editor: Marna Golub-Smith

Technical Reviewers: Shelby J. Haberman and Tim Moses

Abstract

Maintaining score stability is crucial for an ongoing testing program that administers several tests per year over many years. One way to stall the drift of the score scale is to use an equating design with multiple links. In this study, we use the operational and experimental SAT® data collected from 44 administrations to investigate the effect of accumulated equating error in equating conversions and the effect of the use of multiple links in equating. No equating error is directly observed or calculated in the study. Instead, we focus on the behavior of the equating conversions after a series of equatings under the nonequivalent groups with anchor test design and analyze the effect of equating error on conversions. It is observed that the single-link equating conversions drift further away from the operational ones as more equatings are carried out. Analysis of variance is used to decompose the scale score means and the conversion into two major factors: administration month and year for both single- and multiple-link equating results. Seasonality is seen in the data. In addition, the single-link conversions exhibit a certain instability that is not obvious for the operational data. A statistical random walk model is offered to explain the mechanism of scale drift in equating caused by random equating error.

Key words: score stability, equating design, scale drift

Acknowledgments

The authors would like to thank Dr. Shelby Haberman for his discussion and consultation and Songbai Lin for the daunting task of collecting data for the analysis.

1. Introduction

Equating is used to adjust for small differences in test form difficulty so that scores obtained from different forms are interchangeable. When test samples have different ability distributions, equating with an anchor design attempts to tease out differences in performance that are due to the group's ability differences. This equating design is widely used because of its flexibility.

All equating is subject to equating error, either systematic or random. When a series of individual equatings are concatenated over time, there can be shifts in score conversions. Haberman and Dorans (2009) discussed shifts in conversions and sources of variation. They mentioned that accumulated random error is listed as one of the sources that can lead to systematic scale drift. Even under perfect equating conditions, the error can accumulate to some intolerable degree after a series of equatings. A recent study (Guo, 2010) also showed analytically that accumulated equating error will not converge after many equatings. As discussed by Kolen (2006), "even though an equating process can maintain the score scale for some time, the cumulative effects of changes might result in scores at one time being not comparable with scores at a later time" (p. 169). Determining how to maintain score stability is crucial for a testing program that administers several tests per year over many years.

One way to impede the drift of the score scale is to use an equating design with multiple links. Harris and Kolen (1994) examined the stability of equating in the random groups design over a number of links and concluded that using a conversion that was the average across multiple links might be better than using a conversion from individual links. Hanson, Harris, and Kolen (1997) compared single- and multiple-link equipercentile equating with a random groups design using the bootstrap technique (Kolen & Brennan, 2004). It was found that the standard error of equating of the average equating function across links is less than or equal to the standard error of equating for any of the individual links used in the average. Kolen and Brennan (2004) provided an analytical explanation for the reduced error in the average equating function for one equating. Haberman, Guo, Liu, and Dorans (2008) used the SAT® I Reasoning Test as an example to examine the effect of multiple linking in equating. It was observed that the SAT I, which uses four links in each equating under the nonequivalent groups with anchor test design (NEAT), managed to maintain comparable score scales for the 9 years studied (1995–2003).

In this study, we collected SAT® data from 44 administrations to compare single-link equating and multiple-link equating and to explore the effects of accumulated equating error resulting in scale drift after a long series of equatings. We use both operational data and experimental data. The operational data were from multiple-link equatings, and the experimental data were from single-link equatings. No equating error is directly observed or calculated in the study. Instead, we focus on the behavior of the conversions after a series of equatings under the NEAT design and analyze the effect of equating error on conversions.

In section 2, we describe the data collection from the single- and multiple-link equatings. The single-link equating results were created from the operational results. The operational equating results served as a criterion to compare with the newly created single-link equating results. In section 3, analysis of variance (ANOVA) is used to decompose the scale score means and the conversions into two major factors: administration month and year for both single- and multiple-link equating results. Section 4 discusses the results. The statistical random walk model is offered to explain the mechanism of scale drift in equating caused by equating error. Appendix A describes the main features of a random walk.

2. Data

The data used in the study are the scale score means and raw-to-scale conversions for 44 SAT I Verbal and Math forms administered from April 1996 to November 2003. These administrations occurred after the SAT recentering (Dorans, 2002) and before the SAT revision in 2005, a relatively stable period for SAT. Recentering set the scale score mean at 500 and the standard deviation at 110 for the 1990 reference group (Dorans, 2002a, 2002b). In addition, the scale scores were set to be approximately normally distributed in the 1990 reference group.

Operational administrations for the months of March,[1] May, June, October, November, and December were used for the study. In each administration, the SAT Verbal contained 78 items, and the SAT Math contained 60 items. Raw scores are raw formula scores: correct responses received a score of 1, omitted responses and incorrect student-produced responses received a score of 0, incorrect responses to multiple choice questions received a score of $-1/4$ if five choices were presented, and incorrect responses received a score of $-1/3$ if four choices were presented. In creating total raw (formula) scores, the sum of the item scores was rounded to yield an integer value, and raw scores could be negative.

In each operational equating, the new form is equated back to four old forms through four different anchors, respectively. Among the four old forms, one is called the short leg, which was administered 1 year prior, and the other three are the long legs, administered 2 years prior. The ability of the group taking the short leg form is similar to the ability of the group taking the new form because the forms are administered in the same month of the year. For each single link, the raw scores on the new form are equated to the raw scores on the old form (the raw-to-raw conversions) and then mapped to the old-form scale. The table that links the new form raw scores to the scale is called the raw-to-scale conversion. The final/operational raw-to-scale conversion is the weighted average (the average of the short leg and the average of the three long legs) of the four individual conversions. The 44 operational conversions are referred to as our multiple-link data set in the study.

The experimental/new data used in this study were based on equating to short-leg forms only.The single-link equatings were obtained by acting as if each new form from 1996 forward was equated back to the test form administered at the same time of year during the prior year. For example, the 1996 December form was equated to the 1995 December form and placed on the 200-to-800 scale via the raw-to-scale conversion for that December 1995 form. The resultant single-link raw-to-scale conversion for December 1996 was used to place the December 1997 test form on scale, and so forth up to 2003. The single-link raw-to-scale conversions for 1996–2003 for March, May, June, October, November, and December were obtained in this way and are depicted in Figure 1. Independently, we obtained six equating strains, as indicated in Figure 1[2] for March, May, June, October, November, and December. Comparison of operational and experimental data sets is feasible because they are obtained from the same populations, the same new forms, the same equating samples for the short-leg forms, and the same equating designs. The only difference is the number of equating links.

Note that in our study, the scale scores used in raw-to-scale conversions are not the same as the reported scale scores. The scale scores in this study are accurate to four decimal places. Examinees receive reported scale scores that are expressed in integer multiples of 10. Reported scores range from 200 to 800 so that a scale score less than 200 is reported as 200 and a scale score greater than 800 is reported as 800.

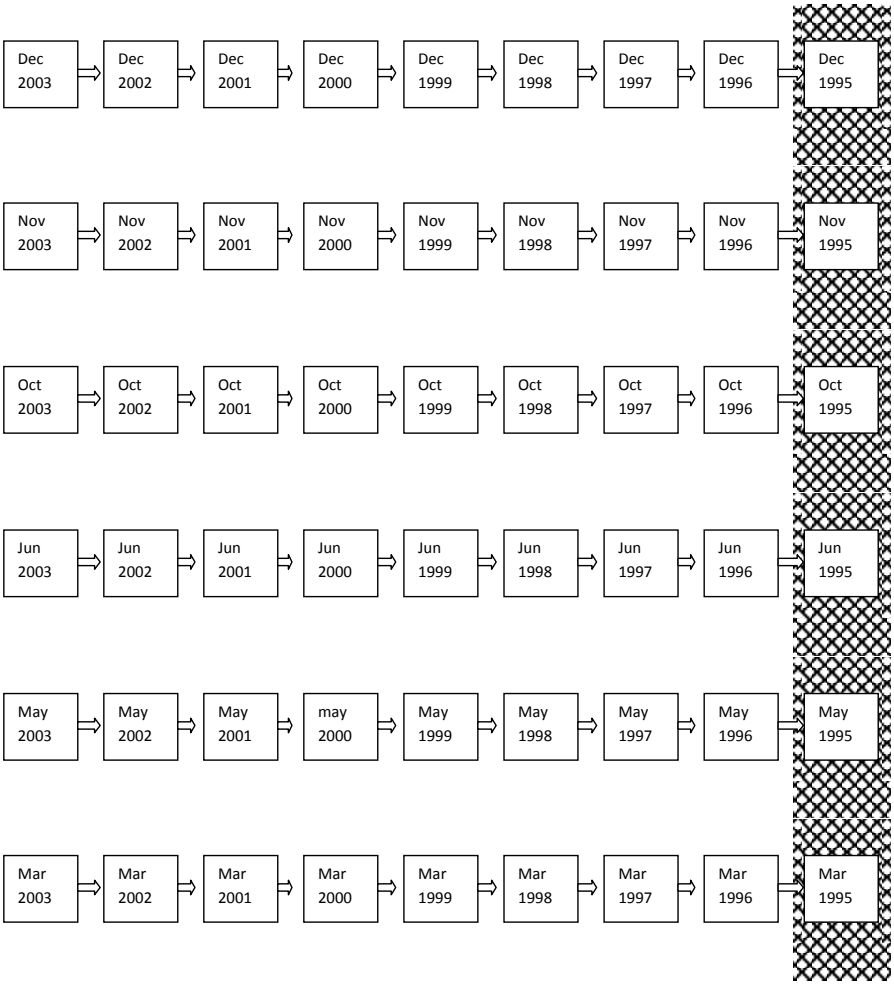

Figure 1. **Equating with a single link.**

3. Results

3.1 Descriptive Statistics

The raw-to-scale conversions are plotted for each administration month. For example, Figure 2 plots the difference between the new and operational raw-to-scale conversions for Verbal. Notice that more recent raw-to-scale conversions include increased numbers of intermediate equatings.

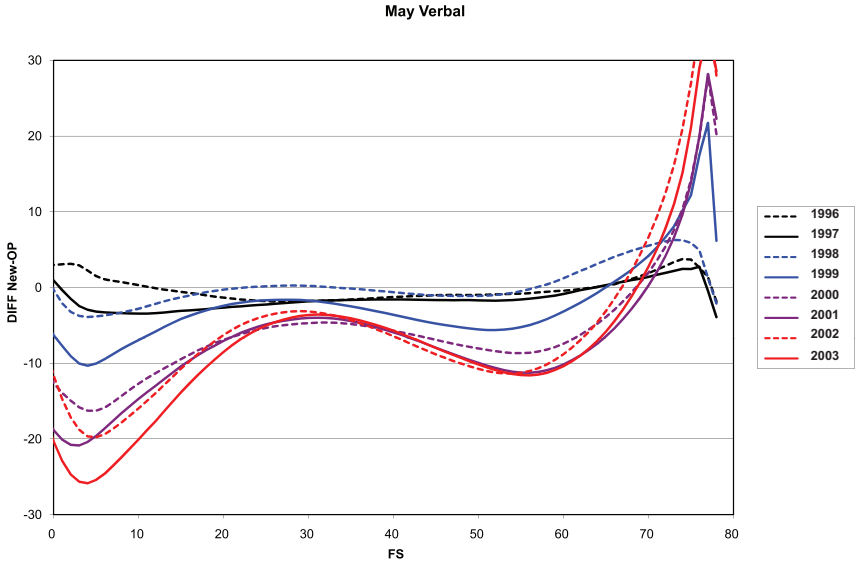

Figure 2. **May conversion: Verbal.**

The x axis (FS) in the plots starts from the raw formula score of zero because scores below zero may be obtained by extrapolation, instead of equating. The y axis (DIFF New-OP) is the difference between the single- and the multiple-link raw-to-scale conversion. Because the reported scores for SAT are integer multiples of 10, any difference in unrounded conversions less than 5 points can be ignored. It is observed that the differences between the single- and multiple-link raw-to-scale conversions tend to get larger, exceeding 5 points at many raw score values, as more equatings are carried out. This pattern is observed for both Verbal and Math conversions for all administration months. The complete set of difference plots (Figures A1–A12) is given in Appendix A.

Scale score means were calculated for each administration for single- and multiple-link data sets. Let V_o, V_n, M_o, and M_n denote the operational scale score mean for Verbal, the newly

created scale score mean for Verbal, the operational scale score mean for Math, and the newly created scale score mean for Math, respectively. Figure 3 plots the difference between V_o and V_n for Verbal for the 44 administrations; Figure 4 plots the difference between M_o and M_n for Math. The difference between the single- and multiple-link scale score means increases across years for both Verbal and Math.

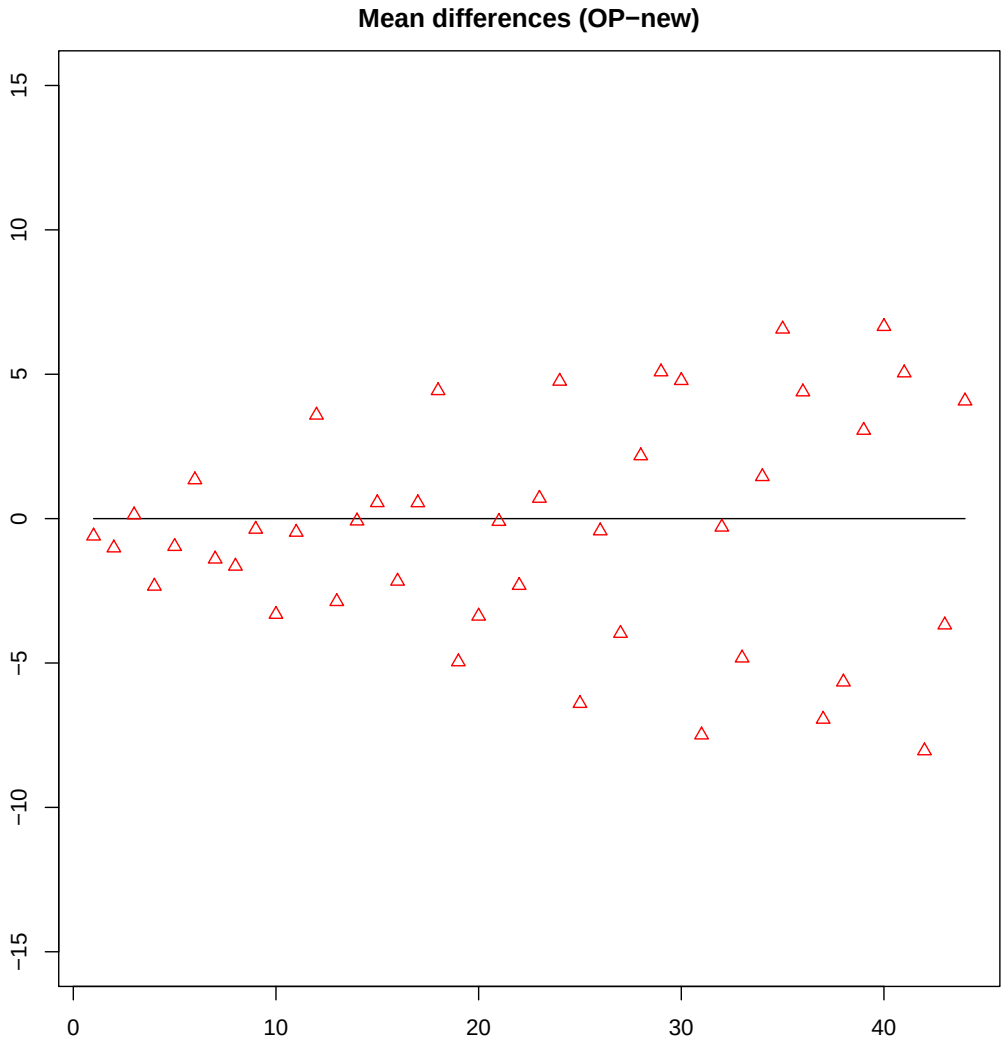

Figure 3. **Mean differences between V_o and V_n for Verbal; $\sigma(V_o) = 11.74$, $\sigma(V_n) = 9.81$. The differences between new and old means increases across years. The x axis stands for the number of administrations from 1 to 44.**

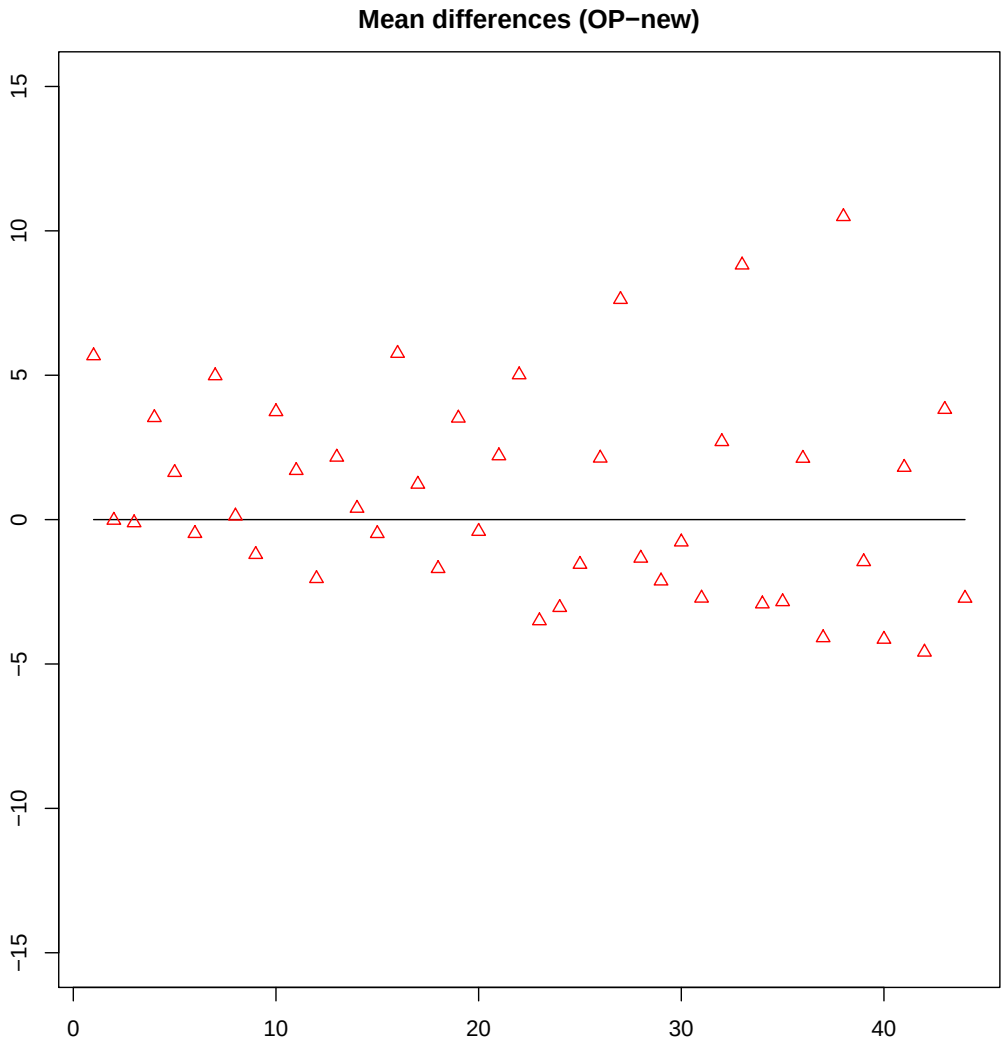

Figure 4. **Mean differences between M_o and M_n for Math; $\sigma(M_n) = 15.46$, $\sigma(M_o) = 13.18$. The difference between the new and old means increases across years. The x axis stands for the number of administrations from 1 to 44.**

3.2 Summary Statistics

As has been observed previously by Haberman et al. (2008), SAT data show strong seasonality. This seasonal variation is also observed in the single-link data (refer to Figure 5). For example, October always tends to have the highest mean and December the lowest. This pattern was repeated for the studied 8 years.

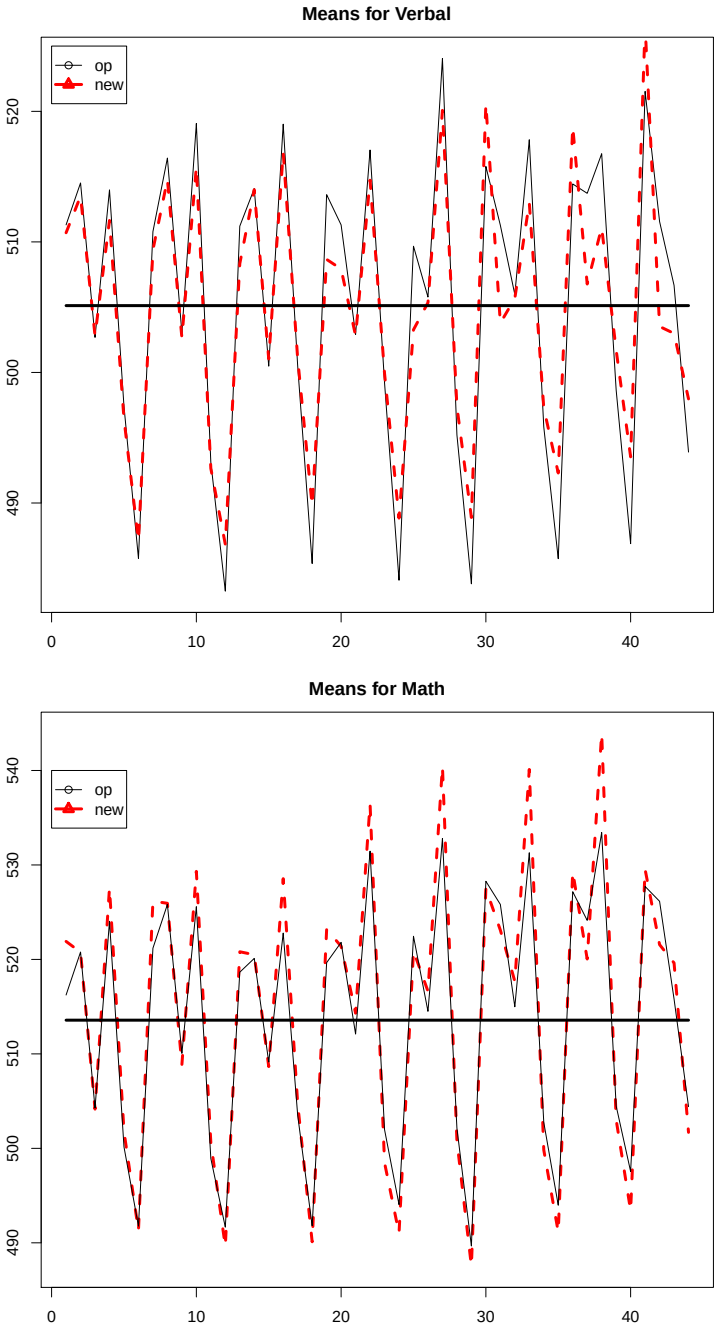

Figure 5. **Score means for Verbal and Math both show seasonality. In the plots, the y axis stands for the score mean, and the x axis stands for the administration number (1–44). The solid horizontal line is the average of all 44 operational score means.**

As discussed by Haberman et al. (2008), the potential factors that contribute to the variation of score means are administration month and year. The adminstration month effect reflects the seasonal variation, and the year effect may indicate the scale drift over time. The same two-way analysis of variance (Haberman et al., 2008) is used here:

$$V_t = \mu + \alpha_{m(t)} + \underline{y}(t) + \gamma_{y(t)} + \delta_{m(t)} y(t) + e_t, \tag{1}$$

with identifiability constraints

$$\sum_i \alpha_i = \sum_j \gamma_j = \sum_j j\gamma_j = \sum_i \delta t a_i = 0, \tag{2}$$

where, in this example, V_t is the Verbal mean at the administration time t. As is customary in ANOVA, the errors e_t are assumed to be independent and to have mean zero and common variance. The $\alpha_{m(t)}$ term corresponds to a month effect. The year effect is $\underline{y}(t) + \gamma_{y(t)}$. The interaction $\delta_{m(t)} y(t)$ is assumed to be linear in the year code $y(t)$.

Tables 1 and 2 present the two-way ANOVAs of the Verbal means V_o and V_n; Tables 3 and 4 present ANOVAs of the Math means M_o and M_n. In this report, we focus on the comparison of the single- and multiple-link equating results. More discussion on ANOVA of similar results can be found in the work of Haberman et al. (2008).

From these tables, administration month is observed to be the main factor in the scale score mean variation for both Verbal and Math. The month factor accounts for about 85%–95% of the total variation in the means. The linear in the year factor (interaction) is also significant, but the effect size (portion of total variation explained by this factor, as shown in parentheses in the "Sum Sq" column in Tables 1–4) is relatively small for both Verbal and Math. The year factor is significant for Math but not for Verbal; however, this factor has a relatively small effect size for Math. In addition, the month factor explains more variation for Verbal in the operational data than in the newly created data. For Math, they are consistent.

The seasonality can also be observed from the left-hand panels of Figures 6 and 7. October populations tend to have the highest ability and December populations the lowest. From the right-hand panels of Figures 6 and 7, the variance of the means are rather stable across years; the operational means are slightly more stable.

Table 1

ANOVA of Operational Verbal Means (V_o)

	Df	Sum Sq (Proportion)	Mean Sq	F value	Pr($>F$)
Month	5	5655.57 (95.5%)	1131.11	188.11	.0000
Year	7	23.21 (0.4%)	3.32	0.55	.7876
Month:yt	5	86.55 (1.5%)	17.31	2.88	.0337
Residuals	26	156.34 (2.6%)	6.01		

Note. $R^2 = 0.9735994$.

Table 2

ANOVA of Newly Created Verbal Means (V_n)

	Df	Sum Sq (Proportion)	Mean Sq	F value	Pr($>F$)
Month	5	3479.14 (84.1%)	695.83	82.68	.0000
Year	7	43.67 (1%)	6.24	0.74	.6395
Month:yt	5	395.72 (9%)	79.14	9.40	.0000
Residuals	26	218.81 (5.3%)	8.42		

Note. $R^2 = 0.9471141$.

Table 3

ANOVA of Operational Math Means (M_o)

	Df	Sum Sq (Proportion)	Mean Sq	F value	Pr($>F$)
Month	5	7000.08 (93.7%)	1400.02	336.98	.0000
Year	7	301.22 (4%)	43.03	10.36	.0000
Month:yt	5	63.36 (0.8%)	12.67	3.05	.0269
Residuals	26	108.02 (1.4%)	4.15		

Note. $R^2 = 0.9855446$.

Table 4

ANOVA of Newly Created Math Means (M_n)

	Df	Sum Sq (Proportion)	Mean Sq	F value	Pr($>F$)
Month	5	9666.39 (94.1%)	1933.28	456.71	.0000
Year	7	212.62 (2.1%)	30.37	7.18	.0001
Month:yt	5	284.76 (2.8%)	56.95	13.45	.0000
Residuals	26	110.06 (1.1%)	4.23		

Note. $R^2 = 0.9892875$.

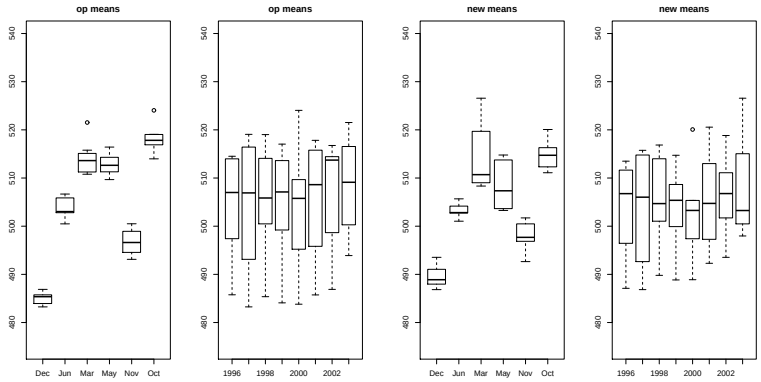

Figure 6. **Box plots of Verbal means by month and year. Outliers, presented as circles, are points outside 3/2 times the interquartile range from Q_1 or Q_3; the whisker are extended to the farthest points that are not outliers. The plot shows seasonality of the means. The operational means are slightly more stable than the new ones across years with regard to the interquartile range.**

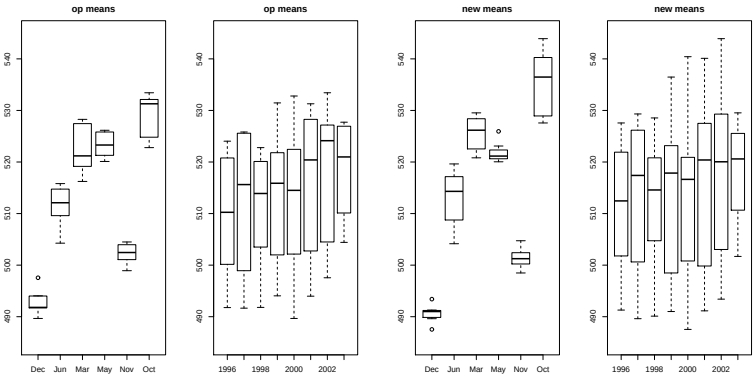

Figure 7. **Same as Figure 6, but for Math means.**

3.3 Raw-to-Scale Conversions

The same ANOVA used for the summary statistics was applied to the raw-to-scale conversions at each raw score point to verify whether the conversions have seasonality and whether the month and year factors play a role in the variation of conversions at each raw score point. Figure 8 displays variance ratios of each component to the total variation in the Verbal conversions for components month, year, linear interaction of month and year, and total model, respectively. If there are no effects of month, year, or interaction, the expected values of the ratios are $5/43 = 0.12$ for month, $7/43 = 0.16$ for year, $5/43 = 0.12$ for interaction linear in year, and $17/43 = 0.40$ for total. For example, in Figure 8 (top left), the y axis is the variance ratio of the month component to the total variation; the x axis is the raw score; the solid line and the dashed line are the variance ratio of month to total for the multiple-link and single-link conversions, respectively; and the dotted line is the expected ratio 0.12. From Figure 8 (top right and bottom left), the year effect and linear-interaction effect of the multiple- and single-link equating conversions are similar, but the single-link conversions are slightly closer to the expected values. From Figure 8 (bottom right), the ratio of the total linear model variation to the total data variation of the multiple-link conversions is much closer to the expected value 0.4 than the single-link conversions. Figure 9 presents the component analysis for the Math conversions. For both Verbal and Math, the ratios of the multiple-link/operational conversions are closer to the expected values. This indicates that multiple-link equating is more stable than single-link equating.

3.4 Means Using Common Weights

Notice that in the summary statistics presented in section 3.2, test form differences and ability differences are confounded. To control the ability difference effect and to provide a summary of the conversions and their differences (exhibited in Figures 12–23) between the single- and multiple-link equatings, we calculate the means of the raw-to-scale conversions using the same weights. By doing so, the seasonal effect will be removed from the means with common weights. For example, should the test forms be parallel to each other, the raw-to-scale conversions will be the same across the 44 administrations. The means with common weights for the 44 administrations will be the same, too, which reflects the consistency of test form difficulties. However, the operational means, or the means obtained using population frequencies at each score level, are population dependent. The variation of the operational means can be attributed to test

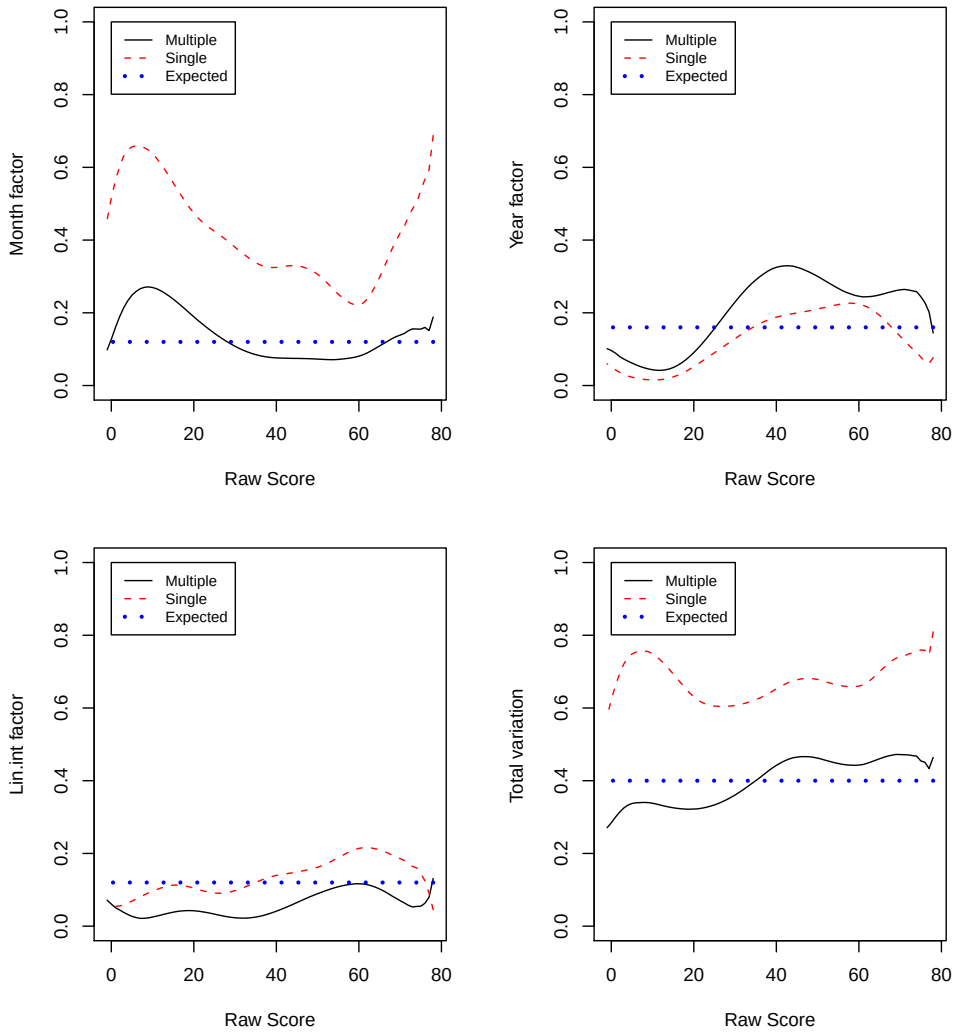

Figure 8. **Component analysis of the raw-to-scale Verbal conversions.**

form variation as well as to population variation.

More specifically, let f_{ij} be the frequency of the jth administration at raw score i, and let s_{ij} be the scale score corresponding to raw score i in the jth raw-to-scale conversion. Define the common weights:

$$w_i = \sum_{j=1}^{T} f_{ij}/T, \tag{3}$$

where T, the total number of administrations, is 44 in our data. Then the means using common weights (MUCW) for the jth administration are $\bar{S}_j = \sum_{i=m}^{M} w_i s_{ij}$, where m and M are the minimum and maximum raw scores in the conversion, respectively. We use $\bar{S}(V)_j$ and

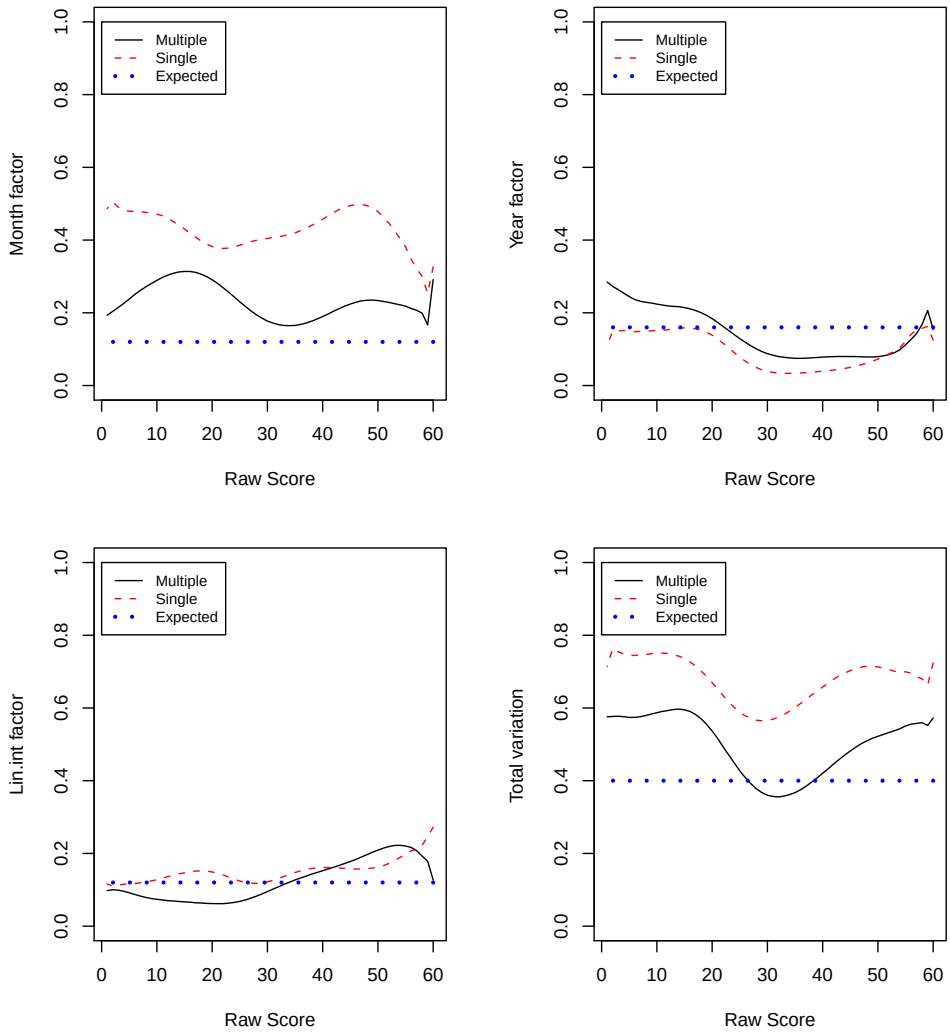

Figure 9. **Component analysis of the raw-to-scale Math conversions.**

$\bar{S}(M)_j, j = 1, \cdots, T$, to denote the MUCW for Verbal and Math, respectively.

Figure 10 contains the box plots of the Verbal MUCW by month and year for the operational data and the single-link data. The operational MUCW and the new single-link conversions behave somewhat similarly with respect to month but not to year. It can be observed in Figure 10 (top left) that the seasonal pattern for the operational MUCW is not as obvious as in Figure 6 because the ability differences for different administration months are removed in the MUCWs. However the MUCWs (Figure 10, bottom left) for the newly created single-link data are not as stable as the operational ones. More importantly, in Figure 10 (bottom right), we observed that the interquartile range (IQR) of the MUCWs for the newly created data has an increasing

trend across years overall, but this is not the case for the MUCWs of the operational data. This trend of increasing variation across years reflects and summarizes the drift of raw-to-scale conversions in section 2. For Math, Figure 11 sends a similar message regarding the Verbal results.

Different weights w_i can be applied to Equation 1, for example, uniform weights $w_i \equiv 1/(M-m)$, to provide the summary of the conversions. But one would expect to obtain the same observation that the IQR of the MUCWs for the single-link equatings increases across years more obviously than the multiple-link equatings; that is, the single-link conversions tend to exhibit more variation across years than the multiple-link conversions.

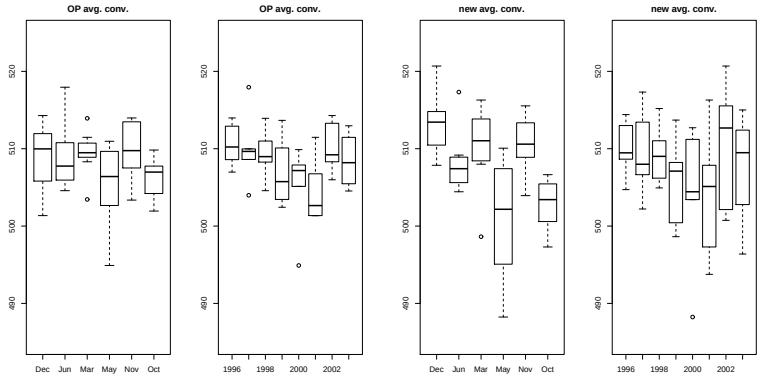

Figure 10. **Box plots of Verbal MUCWs. Overall, IQRs of the new MUCWs are increasing across years.**

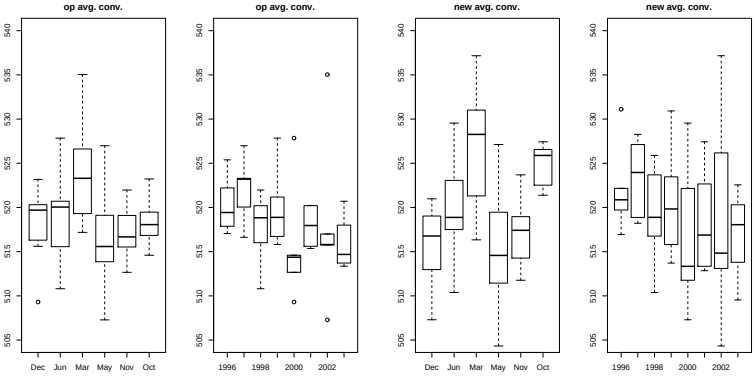

Figure 11. **Same as Figure 10, but for Math MUCWs.**

4. Discussion

In this study, data from 44 SAT administrations were used to produce multiple- and single-link equating results. The single-link equating results showed increased variability in the applied conversions, and they drifted away from the multiple-link (operational) equating results over time. In contrast, the operational conversions were more stable and had less variation across administration years.

For the single-link equatings in our data, the old and new group abilities were similar. Thus these data most likely would show small amount of the equating bias that is associated with ability differences between equating samples. However, random error exists even for unbiased equating. Accumulation of random noise forms a random walk (see Appendix B) that leads to increased variation and a tendency to drift away from where it starts after many equatings.

In the multiple-link equating, the accumulative equating error is the average of several random walks. This slows down the process of drift and stabilizes the equating results, given other factors, such as test populations, equating sample sizes, and equating designs that are the same. Furthermore, having more than one link for a new form safeguards against a problematic equating when one link is found to be inappropriate. This allows for a check of the stability of the equating results by comparing the similarity of the conversions produced by the separate links. It may help reduce the effect of bias in one of the links (Hanson et al., 1997).

Equating drift is always a problem as more equatings are done. Our study suggests that using a multiple-link equating design dampens the effects of the equating error accumulation process. How many equating links are enough to produce a satisfactorily stable equating result remains a question for further study.

References

Dorans, N. (2002). Recentering and realigning the SAT score distributions: How and why? *Journal of Educational Measurement, 39*, 59–84.

Guo, H. (2010). Accumulative equating error after a chain of linear equatings. *Psychometrika, 75*, 438–453.

Haberman, S., & Dorans, N. (2009). *Scale consistency.* Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego, CA.

Haberman, S., Guo, H., Liu, J., & Dorans, N. J. (2008). *Consistency of SAT reasoning score conversions* (ETS Research Report No. ETS-RR-08-67). Princeton, NJ: ETS.

Hanson, B., Harris, D., & Kolen, M. (1997). *A comparison of single- and multiple-linking in equipercentile equating with random groups.* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Harris, D., & Kolen, M. (1994). *Stability checks in random groups equating.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Kolen, M. (2006). Scaling and norming. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 155–185). Westport, CT: Praeger.

Kolen, M., & Brennan, R. (2004). *Test equating, scaling, and linking* (2nd ed.). New York, NY: Springer.

Notes

[1] The first administration of a year in which a new form is given is either in March or April.

[2] The data for four administrations were not available for our analysis. However, for convenience, we still show 48 administrations in the diagram.

Appendix A

Conversion Difference Plots

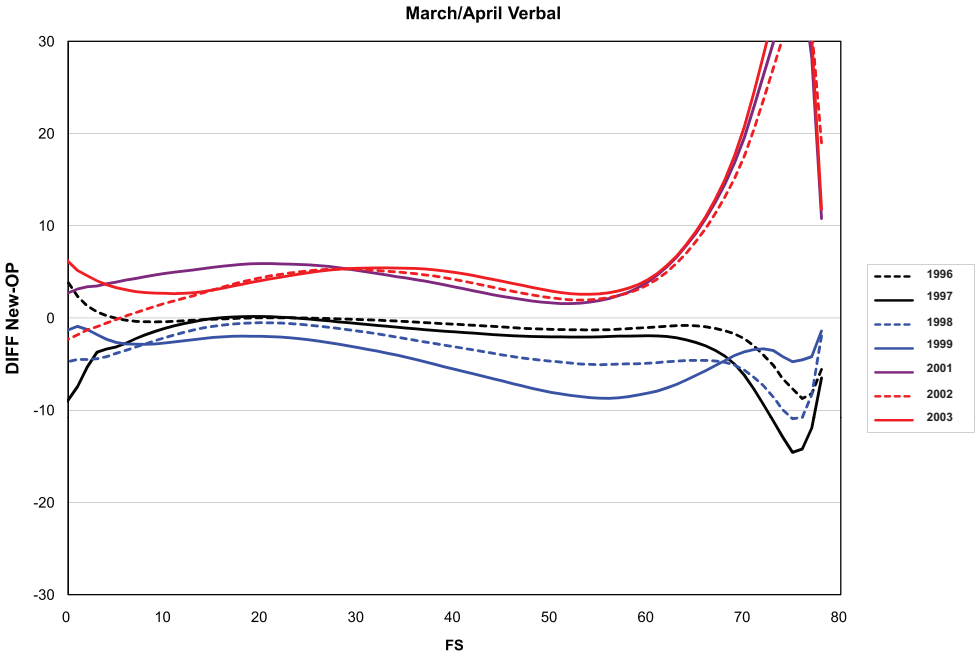

Figure A1. **March/April conversion: Verbal.**

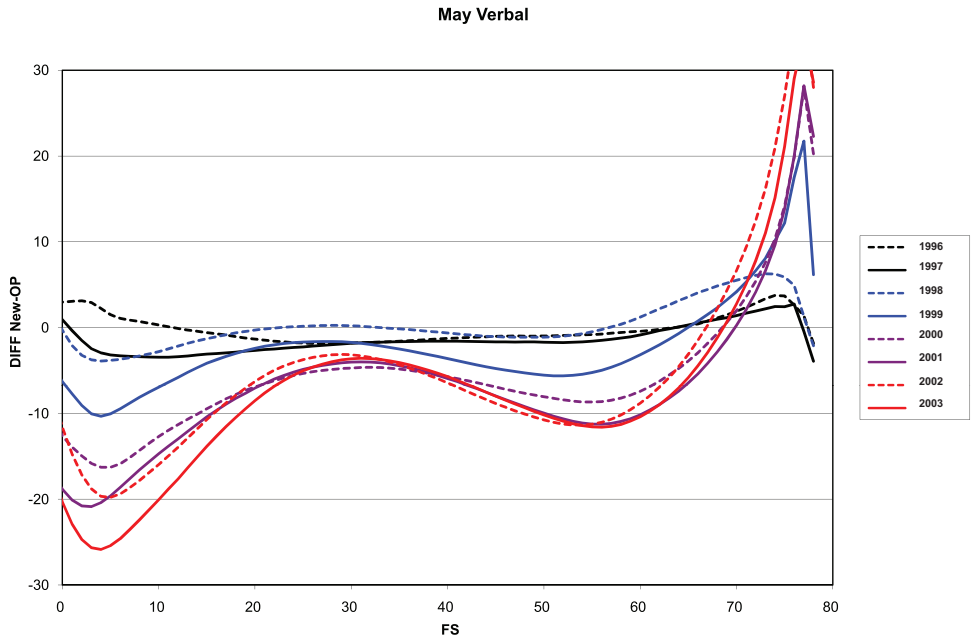

Figure A2. **May conversion: Verbal.**

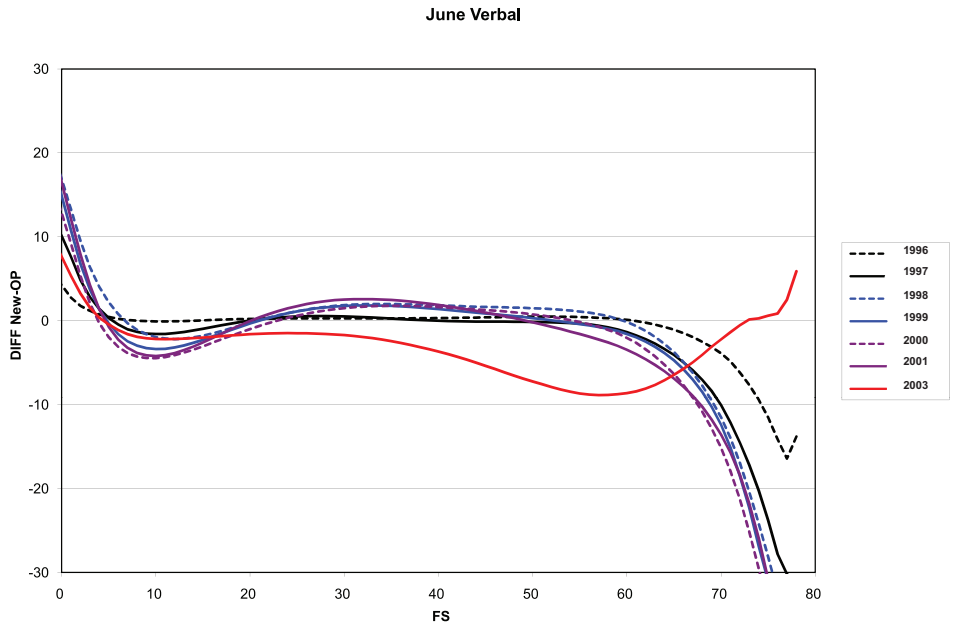

Figure A3. **June conversion: Verbal.**

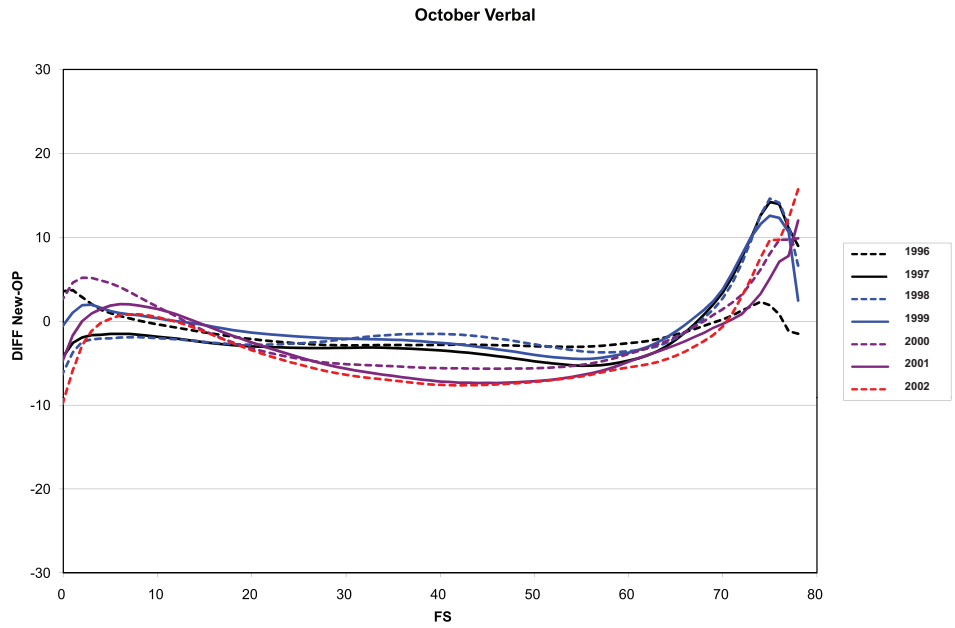

Figure A4. **October conversion: Verbal.**

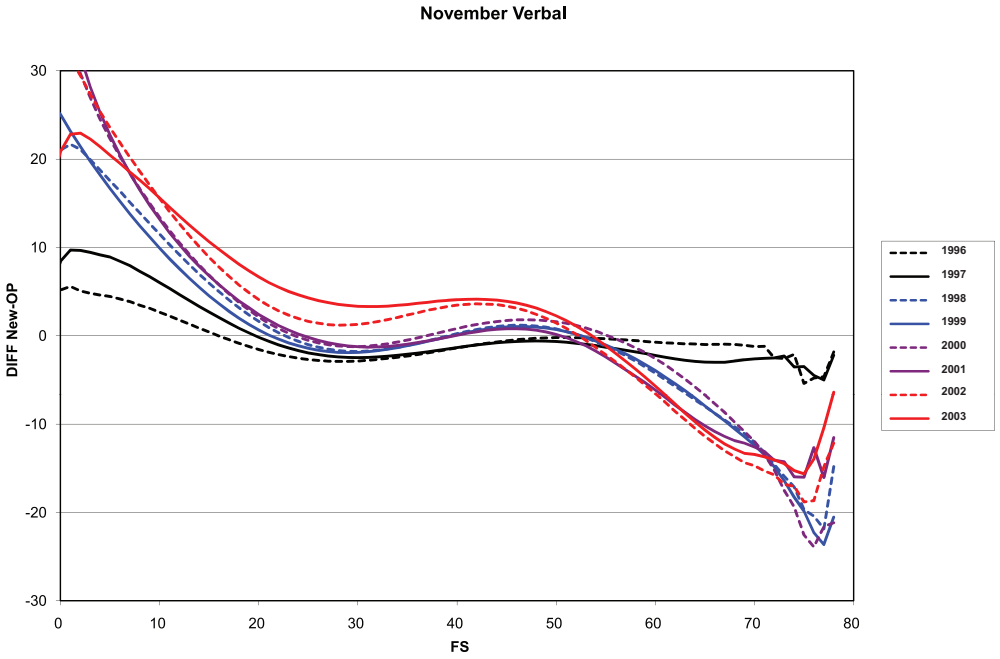

Figure A5. **November conversion: Verbal.**

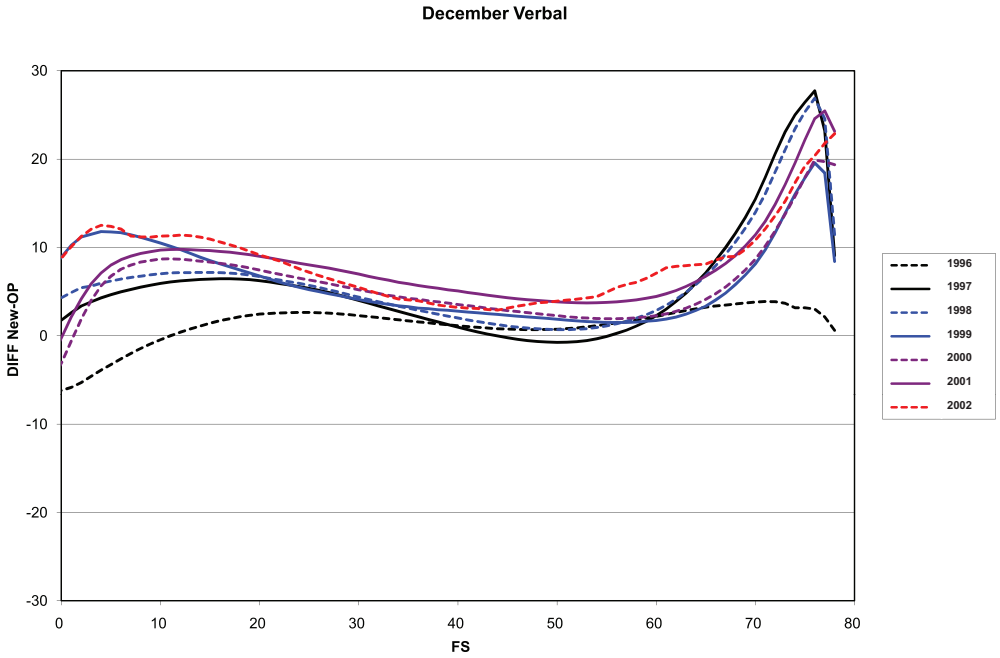

Figure A6. **December conversion: Verbal.**

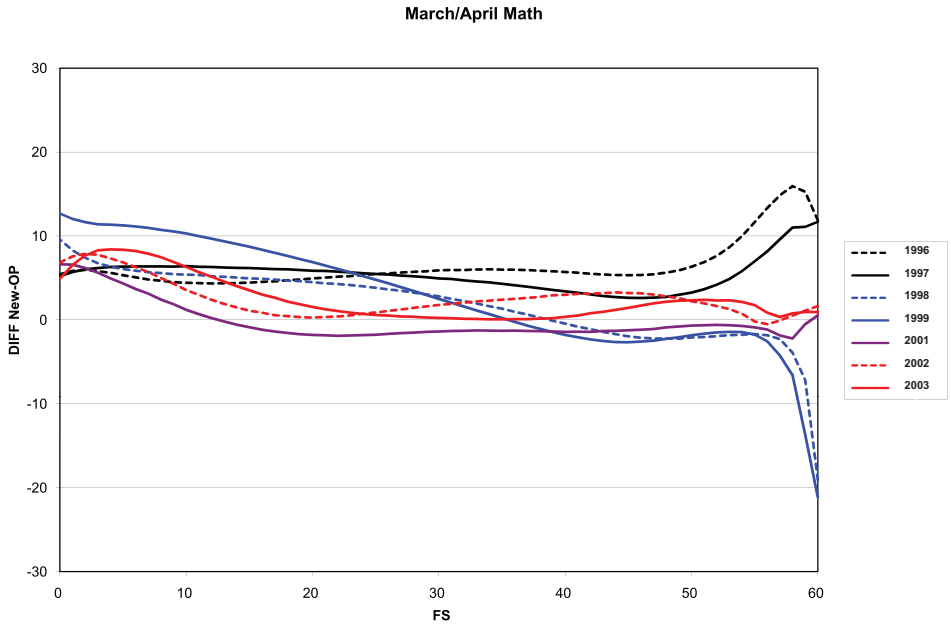

Figure A7. **March/April conversion: Math.**

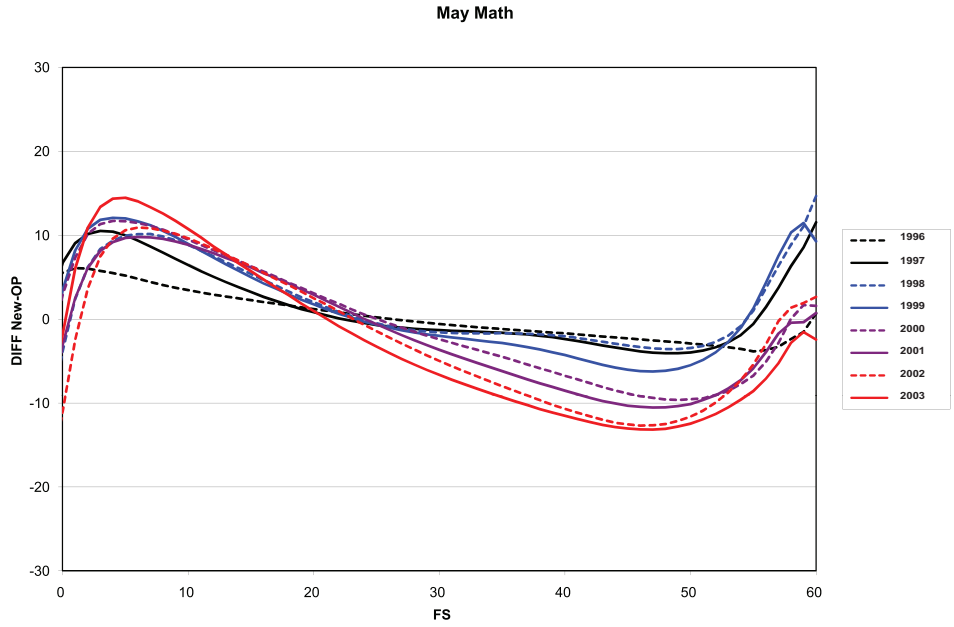

Figure A8. **May conversion: Math.**

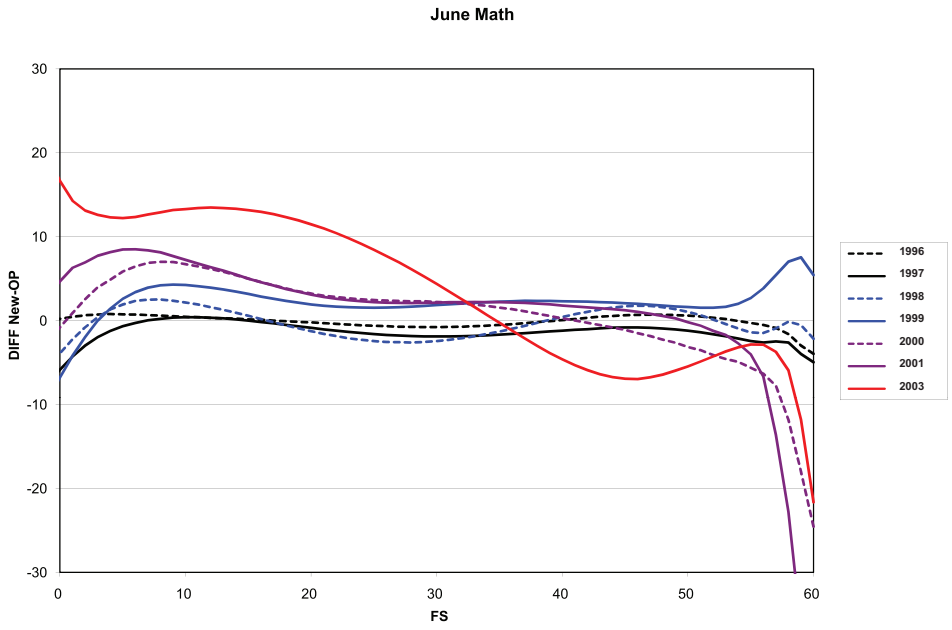

Figure A9. **June conversion: Math.**

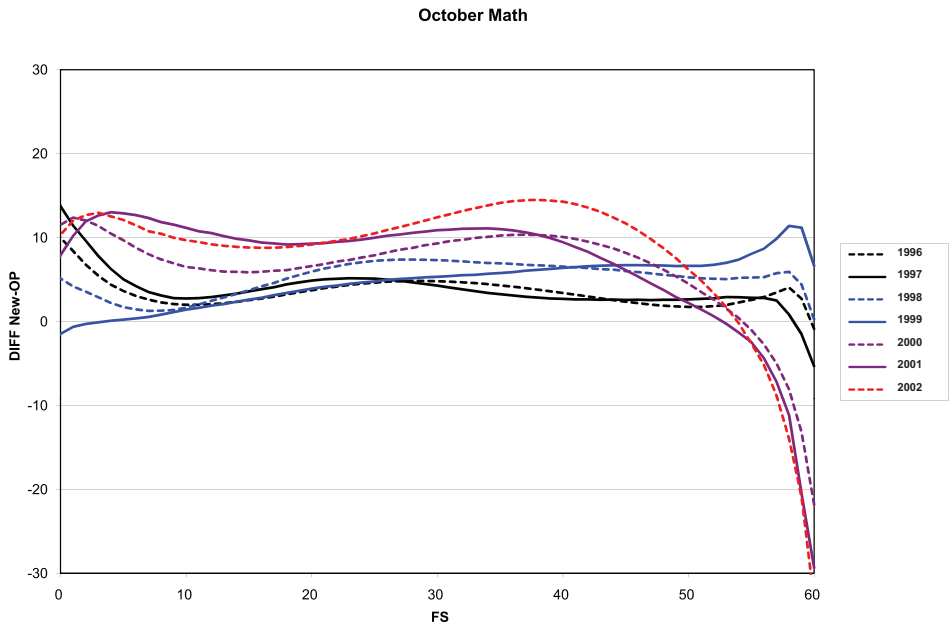

Figure A10. **October conversion: Math.**

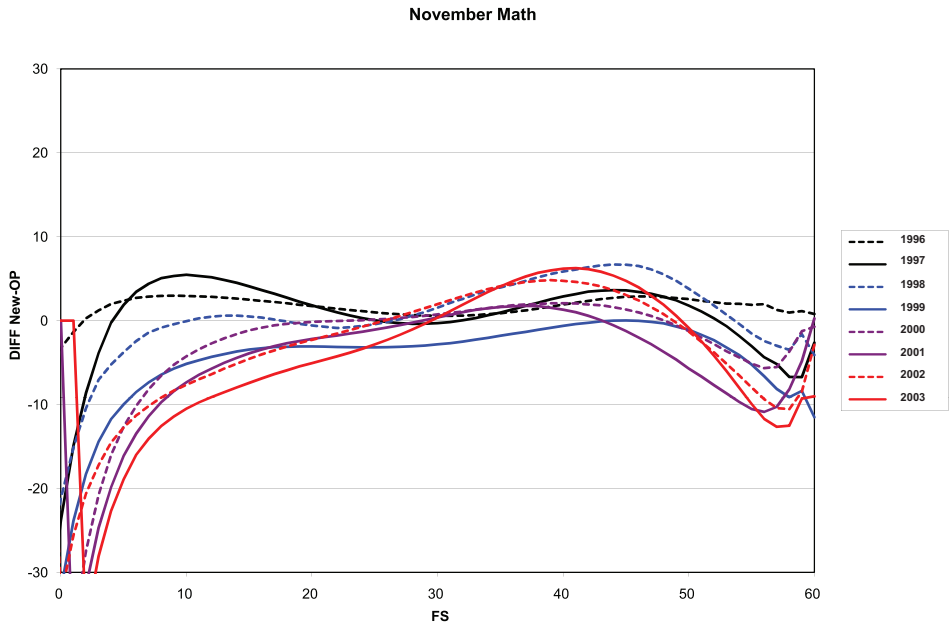

Figure A11. **November conversion: Math.**

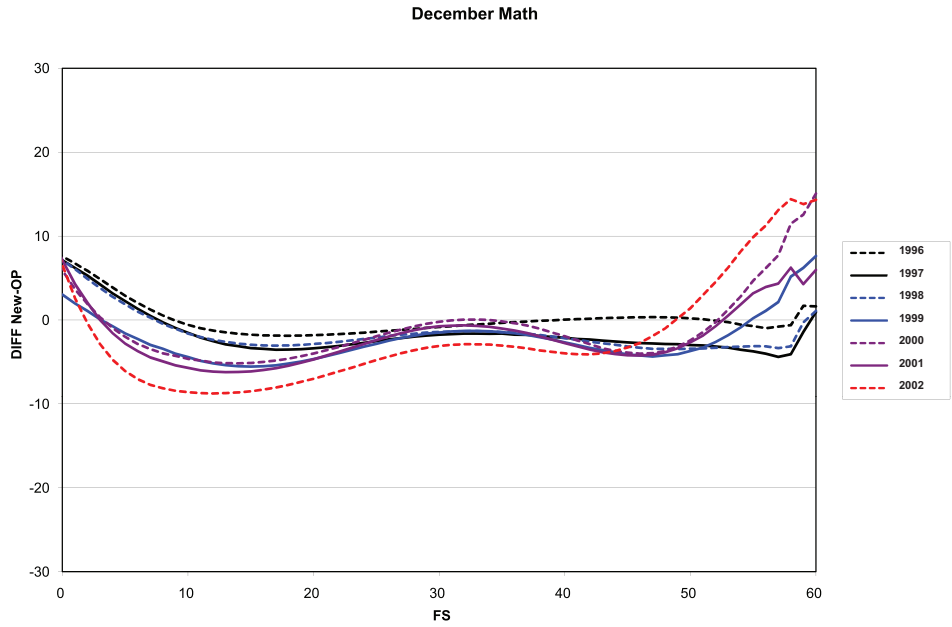

Figure A12. **December conversion: Math.**

Appendix B

Random Walk

To connect what was observed from the SAT single- and multiple-link equating conversions and random walk, we introduce random walk and its main properties. A simple random walk is formed when small quantities of noise are added up. Let $X_t, t = 1, \cdots, n$ be identically and independently distributed random variables with mean zero and variance σ^2. A process Y_t is said to be the simple random walk if $Y_t = Y_{t-1} + X_t$. Usually, $Y_1 = X_1$. Therefore $Y_t = \sum_{i=1}^{t} X_i$. By plotting the position, Y_t, against time in a graph, a representation of the time path of the process is obtained. One fundamental question is what the time path of a random walk looks like. It would be natural to expect the path to scatter randomly around the baseline $Y = 0$ because $\mu(Y_n) \equiv 0$. However, that is not the case. Figure B1 displays the paths of several random walks with different time lengths, where X_t is the standard normal variable. We first focus on the solid lines in the three plots. The solid lines are paths of a random walk for time lengths 50, 100, and 500, respectively.

Following Equation 4, one immediately obtains

$$\mathrm{Var}(Y_n) = n\sigma^2. \tag{B1}$$

This implies that the variation of a random walk is constantly increasing over time. The correlation between the two adjacent states, Y_{n-1} and Y_n,

$$\mathrm{Cov}(Y_n, Y_{n-1}) = \sqrt{\frac{n-1}{n}}, \tag{B2}$$

is approaching unity as $n \to \infty$. This strong correlation explains the relatively stable positions for neighboring states.

Averaging two or more random walks can slow down the explosion of the process. For simplicity, let Y_{1t} and Y_{2t} be two independent simple random walks, and let $Z_t = (Y_{1t} + Y_{2t})/2$. Then

$$\mathrm{Var}(Z_t) = \frac{n}{2}\sigma^2. \tag{B3}$$

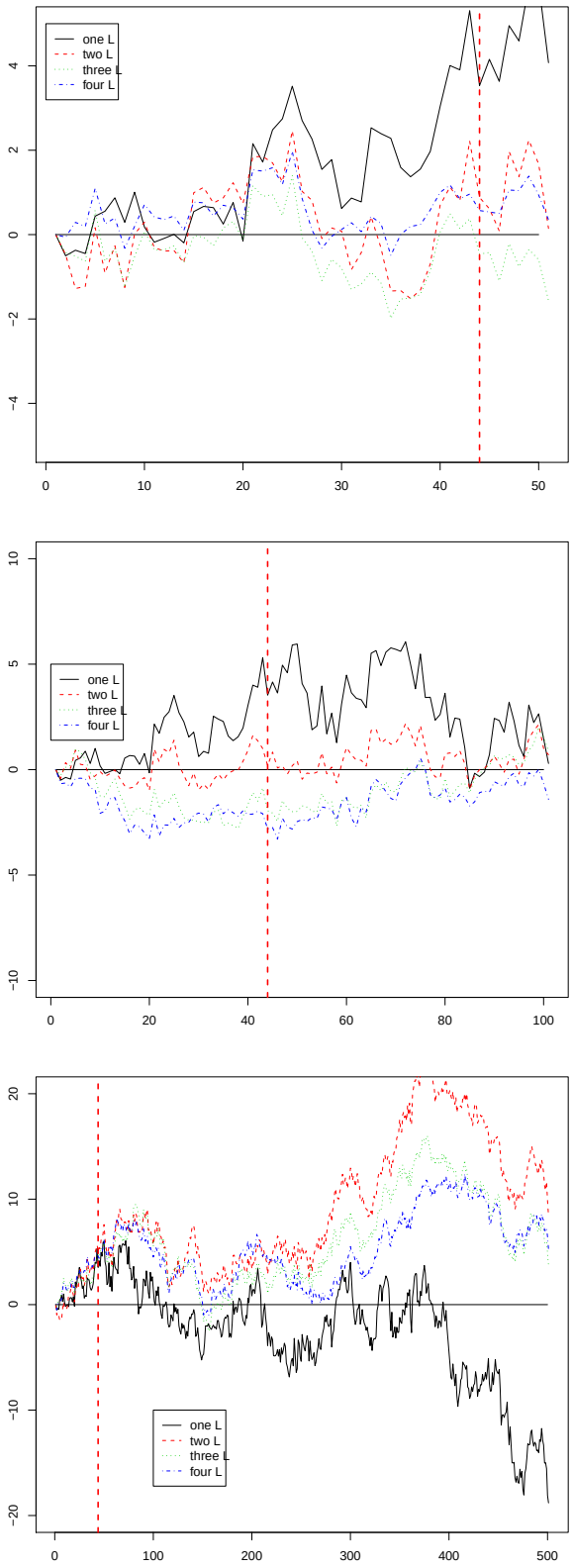

Figure B1. **Averages of random walks.**

In Figure B1, the dashed lines, dotted lines, and the dash-dotted lines are the averages of two, three, and four random walks for different lengths, respectively. The average of a number of random walks seems more stable.

The vertical dashed line at 44 in Figure B1 corresponds to the number of administrations of the SAT in our data set. Random equating error is usually different for different score points; however, as observed in Figures A1–A12, the single-link equating conversions tend to drift away gradually from the multiple-link equating conversions and linger there. This was observed for all equating strains and for both Verbal and Math at different score points, which reflects the strong correlations between conversions at adjacent administrations. The overall variation of the single-link conversions is also observed in Figures 10 and 11 to be increasing over time compared with the multiple-link conversions, which reflects the ever-increasing variation over time. The relative stability of multiple-link equating results is also reflected in Figures 8 and 9. Our data and analysis seem to support the simple random walk model to explain the behavior of single- and multiple-link conversions.