Investigating Displacement in the

Winsteps Rasch Calibration Application

John A. Stahl

Timothy Muckle

Pearson VUE, Evanston, Illinois

Paper presented at the annual

American Educational Research Association Meeting

Chicago, Illinois 2007

Investigating Displacement in the Winsteps Rasch Calibration Application

Introduction:

Item drift analyses that use displacement have reported displacement distributions symmetrically distributed around zero (Jones and Smith, 2006). Approximately equal numbers of items appear to drift in both a positive (harder) and a negative (easier) direction, despite hypotheses suggesting systematic drift in one direction. The possibility exists that a portion of this displacement distribution may be due to a statistical artifact resulting from the way the statistic is calculated. Such an artifact would result in items exhibiting drift when actually their item characteristics have remained stable. The degree that such a statistical artifact affects the drift calculation has not been systematically investigated and its impact on the interpretation of the displacement statistic has not been determined. The purpose of this study is to systematically investigate whether the displacement statistic, as calculated by the Winsteps Rasch calibration program (Linacre 2005), is reflective of actual item drift conditions or if it contains an element of statistical artifact.

The Displacement Statistic:

Displacement is a useful statistic generated from the Winsteps analysis program. The displacement statistic "approximates the displacement of the estimate away from the statistically better value which would result from the best fit of your data to the model." (Linacre, 2005:260). Mathematically, displacement is denoted:

$$\delta = \beta_c - \beta_a \qquad (0.1)$$

Where δ is the displacement statistic, β_c is the freely estimated calibration arising from the current data set, and β_a is the anchor item difficulty. In any analysis featuring anchored items, Winsteps simultaneously performs a free (unanchored) parameter estimation for all of the items. The displacement statistic results from a direct comparison of the anchored difficulty value with the value from the free estimation arising from the current data. Readers are referred to Linacre (2005) and Wright & Stone (1979) for technical details on parameter estimation in Winsteps. Due to the re-centering procedures in Winsteps, the free parameter estimates are constrained to be centered around a mean of 0. Accordingly, all displacement values also sum to 0. As a result, in a dataset featuring systematic drift in one direction (i.e. easier), it is possible to observe, in stable items, drift in the opposite direction (i.e. harder).

In many instances, such as equating tests or administering computer adaptive tests, the difficulties of items are anchored to values derived from previous analyses. When data from subsequent analyses are analyzed and items are anchored to the previously estimated difficulties, then the displacement statistic can be used to monitor the appropriateness of the anchored values and indicate the degree of item difficulty drift. Thus, the displacement statistic is interpreted by a large number of psychometric practitioners and test developers as the logit difference between anchored difficulty value and the difficulty value at which the item is observed in the current analysis. In this way, the displacement statistic is very useful in detecting and monitoring item drift, defined as a systematic change in item difficulty over time.

However, if the displacement statistic contains an artifact that does not represent actual item difficulty drift, then the interpretation of the statistic becomes problematic and its usefulness is diminished. The goal of this study is to use simulation data to replicate certain conditions of item difficulty drift and then to assess the impact of these drift conditions on the displacement statistic and its interpretation.

Methods and Data:

This study used simulation data to represent varying degrees of item drift. The candidate sample was chosen to have a standard normal distribution N(0,1). Three candidate samples were selected, one having 200 individuals, one having 500 individuals and one having 1,000 individuals. These samples were felt to be reflective of typical test sample sizes in the certification and licensure field. Three item samples were also selected, each having a standard normal item difficulty distribution N(0,1). The item sample sizes were 30 items, 100 items and 200 items. These samples again were felt to be reflective of typical test lengths.

Each candidate sample size was then matched against each item sample size, resulting in nine combinations. The Promissor simulator (Becker 2006) was used to generate an initial response string data sample for each of these combinations using Rasch probability as the basis for assigning a right/wrong response for each candidate/item interaction. The nine response strings generated by the above procedure were analyzed using the Winsteps program, and item difficulty calibrations were obtained for each item as it was used in each of the nine combinations. These item calibrations were then used to create item anchor files to be used in subsequent drift analyses.

In order to minimize the impact of outside variation, the same response string data sample was used to simulate item drift. Answers were systematically changed in the response strings to simulate a drift in an easier and/or harder direction by changing answers from wrong to right or vice versa.

The number of items that were simulated to drift was systematically varied. In the first condition, 10% of the items were simulated to drift. In the second condition 20% of the items were simulated to drift. In the final, most extreme condition, 50% of the items were simulated to drift. The first two conditions are probably more reflective of normal drift conditions. The final condition would be more reflective of a serious security breach.

Within each of the above conditions the direction of drift was also varied. In one situation, the drift was all in a single direction with the items becoming easier. This unidirection of drift would simulate a condition in which some items on a test had been compromised and a particular group of candidates had prior knowledge of the items. In the second situation, the drift was symmetrically balanced with half the items drifting easier and half the items drifting harder. This condition is one that might be more typical of normal testing conditions. In the final situation, the drift was asymmetrical with 70% of the items drifting easier and 30% of the items drifting harder. The final condition is a combination of the first two with more emphasis on the condition of compromised items.

Combining the 9 candidate/item combination with the 3 percentages of drift and the 3 directions of drift resulted in 81 unique conditions that were simulated.

To simulate drift, the desired number of items and candidates was randomly selected without replacement from the total candidate and item samples. Each response string was examined and, for each interaction of a selected candidate and a selected item, the answer was examined and changed appropriately to simulate the desired drift. If the answer for that particular candidate/item interaction was already in the direction of the desired drift no action was taken.

The modified data sets were then reanalyzed using Winsteps. The item difficulties of all of the items were anchored to the item calibrations obtained from the initial analysis. The impact of the drift on the displacement distribution was then assessed.

Results:

The results of the simulations and analyses are summarized in Tables 1 and 2. Tables 1 and 2 contain average displacement values observed on items whose response strings were not modified to mimic drift (i.e. items hypothesized to remain stable over time). The mean displacement values are replicated for each condition of the simulations. Table 1 represents simulation cases where hypothetically drifting items were all displacing exclusively in a negative (easier) direction.

Table 1 also demonstrates that when systematic drift in one direction was present in a data set, then hypothetically stable items in every test condition exhibited artificial positive displacement. The artifact was more pronounced in conditions with increased test length and increased proportion of drifting items. The amount of artificial positive drift appeared to be unrelated to examinee sample size. To provide some indication of significance, we compared the Table 1 displacements to the value corresponding to two times the average standard error (SE) of item calibrations for each examinee sample condition (for N=200, 2*SE = 0.34; N=500, 0.22; N=1000, 0.14). Highlighted cells in Table 1 reflect displacement values that might be interpreted as significant because they are more then more than two SE from the original calibration. Obviously, these potential false positives occur exclusively in simulation conditions where 50% of the test items were modified to show easier drift.

Table 2 summarizes simulation cases where hypothetically drifting items were displacing asymmetrically (70% easier, 30% harder).

The patterns of average displacement for hypothetically stable items (i.e. items whose response strings underwent no modification) are similar in Table 2, yet not as pronounced. The artificial positive displacement is detected more in data sets with a large amount of systematic drift. Again, artifact that could be interpreted as significant is highlighted. The problem posed by artifact is somewhat ameliorated in the simulations where all of the drift was not in one direction.

Not surprisingly, in simulations featuring balanced drift – or equal amounts of drift in both directions – the problem of artifact was completely ameliorated. In these simulations, the average displacement value for items hypothesized to remain stable was consistently 0, so their values are not tabulated.

The results of the analyses are also presented graphically in Appendices A through I. Each Appendix presents the results from a particular combination of items and candidates. Thus, for example, Appendix A deals with the results from the 30 item/100 candidates combination.

The various percentage of drifted items and the direction of drift combinations are presented within each appendix. In each of the figures, the horizontal axis is the sequence number of the item on the test. The vertical axis represents the displacement reported by Winsteps. The dark horizontal lines in the figures represent bands of two standard errors centered around zero. The average SE of the item calibration for the sample size of candidates being examined was used in setting these bands.

There are nine figures in each of the appendices. The labels at the bottom of the figure specify the simulation conditions that produced the graphed results. The first three figures display the results from simulating a drift all in an easier direction. The next three figures display the results from simulating half of the items drifting in an easier direction and half of the items drifting in a harder direction. The final three figures display the results from the asymmetrical split of the items, with 70% drifting in an easier direction and 30% drifting in a harder direction.

The effect of the procedure used to calculate the displacement is consistent throughout the various conditions examined. The effect is most easily seen in the results in Appendix A, but repeated in all of the other appendices. The displacement is mean centered around zero which, in turn, forces the total sum of all of the displacement values to be equal to zero.

The most radical situation is where all of the items are systematically drifting in the same directions. In such situations we would expect the drifted items to display a noticeable negative displacement and that is what we see in Figures A-1 and A-2. However, because the sum of the displacement is forced to equal zero, we begin to see the remainder of the items displaying a positive displacement to counter the negative displacement of the drifted items. This shows up a very small value in A-1 since it is spread across all of the non-drifted items, but begins to show a more noticeable value in Figure A-2. By the time we reach the more extreme condition displayed in Figure A-3, the positive displacement of the non-drifted items has increased to the point where it looks to be statistically and substantially significant.

On a more reassuring note, in the situations in which the drift was simulated to be symmetrically distributed in both a harder and easier direction, the affect on non-drifted items is balanced out and they tend to cluster around the zero value. This can be seen in Figures A-4 through A-6.

This balancing-out effect also shows up in the situations where the item drift is asymmetrically distributed as can be seen in Figures A-7 and A-8. However, two areas of concern are associated with this type of condition. The first concern is that the displacement exhibited by the items that are members of the set that has fewer members, in this case the items that are drifting harder, appear to be inflated in value since their sum has to be equal but opposite in sign to the sum of the set that has more members, in this case the items that are drifting easier. This can be noted in Figure A-9.

The second area of concern is associated with asymmetrical distributions of drift where the number of items affected is relatively large. As the number of items drifting in a particular direction increases, it requires a corresponding adjustment in the displacement values of both the items drifting in the other direction and the non-drifting items. This adjustment can be extreme enough so that even the non-drifting items appear to have significantly drifted. This can be seen in Figure 9 in Appendix C and the following appendices.

The patterns discussed above were repeated in all of the appendices in a consistent manner. The areas of concern all appear to arise from the fact that the displacement is forced to be mean centered.

The situations in which items have a displacement that are due to a statistical artifact are relatively easy to detect. An examination of Figure 3 in each of the appendices shows that, in situations in which the artifact most clearly manifests itself, the variability in the resulting displacement is very low, particularly compared to the variability of the displacement of the drifted items. A plot of displacement against sequence number appears to be a very quick and easy visual way of identifying potential artificial displacements.

Discussion:

The displacement statistic is a very valuable tool in monitoring item drift conditions. However, if some of the reported values are impacted by a statistical artifact component, this can have a major influence on how this statistic is interpreted and on the usefulness of these interpretations. A better understanding of the displacement distribution will lead to improvements in the interpretation of this statistic.

The statistical artifact is a result of the way the displacement statistic is calculated and the fact that the results are mean centered. In situations in which the degree of drift is symmetrically distributed in both an easier and a harder direction, the impact is relatively minor. As the drift becomes more asymmetrically distributed, the impact of the artifact becomes more noticeable to the extent that non-drifted items can be flagged as significantly and substantially drifted.

However, the impact can be easily detected by plotting the displacement against the sequence number. Items with a very consistent drift are items that are being affected by the artifact. It is recommended that such a plot be used as a part of any drift analysis. It may also be possible to detect the artifact during the displacement calculation by determining the variability of the drift within subsets. A subset that exhibits little variability may reflect artifact.

Future research will also address possible procedures to modify the drift calculation in a way to remove the artifact. The drift simulation was straight forward in this study. Future study of this topic will also use a more probabilistic simulation of drift to mimic item displacement on a more realistic basis.

References

Becker, Kirk (2006) Promissor CAT Simulator

Jones, P.E. and Russell W. Smith (2006) *Item Parameter Drift in Certification Exams and Its Impact on Pass-Fail Decision Making*, Paper presented at the annual meeting of the National Council of Measurement in Education, San Francisco.

Linacre, John M. (2005) *A Users Guide to Winsteps.*

Table 1

Mean displacement values for hypothetically stable items (unidirectional drift)

Condition	Examinee Sample	30 item test	100 item test	200 item test
10% modified, all easier	N=200	0.01	0.03	0.03
	N=500	0.02	0.02	0.02
	N=1000	0.03	0.02	0.03
20% modified, all easier	N=200	0.11	0.12	0.13
	N=500	0.09	0.11	0.10
	N=1000	0.12	0.09	0.10
50% modified, all easier	N=200	0.66	0.70	0.77
	N=500	0.55	0.72	0.74
	N=1000	0.61	0.78	0.68

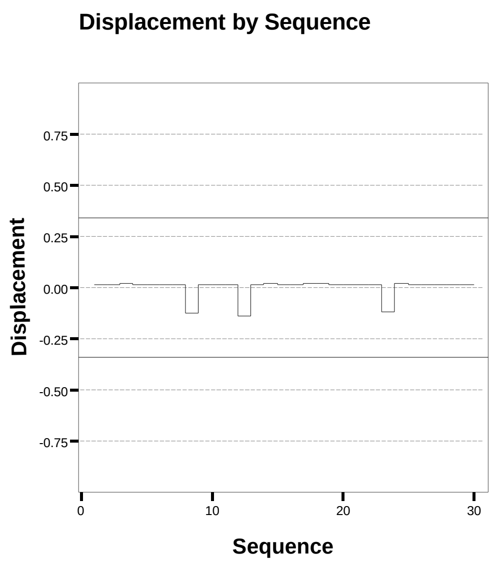
Table 2

Mean displacement values for hypothetically stable items (asymmetrical drift)

Condition	Examinee Sample	30 item test	100 item test	200 item test
10% modified	N=200	0.00	0.02	0.01
	N=500	0.00	0.01	0.01
	N=1000	0.01	0.01	0.01
20% modified	N=200	0.04	0.05	0.05
	N=500	0.03	0.06	0.04
	N=1000	0.02	0.03	0.03
50% modified	N=200	0.12	0.29	0.39
	N=500	0.00	0.30	0.25
	N=1000	0.12	0.21	0.29

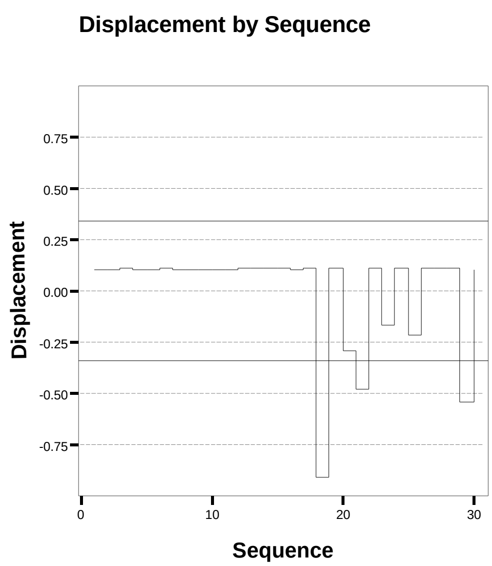
Appendix A: 30 item vs. 200 candidate combination

Figure A-1



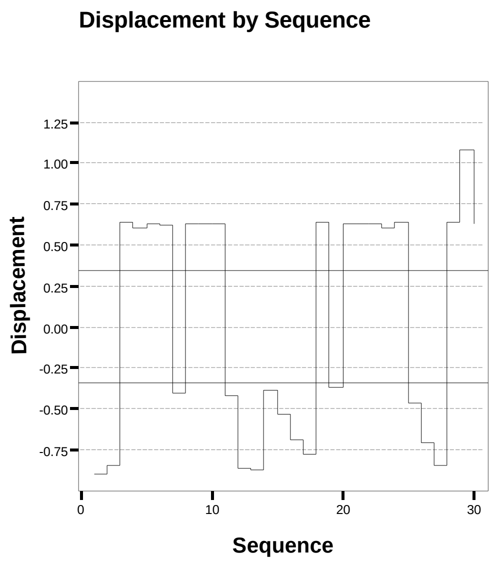
30 item test with 200 candidates
10 percent of the items drifted easy
for 10 percent of the candidates

Figure A-2



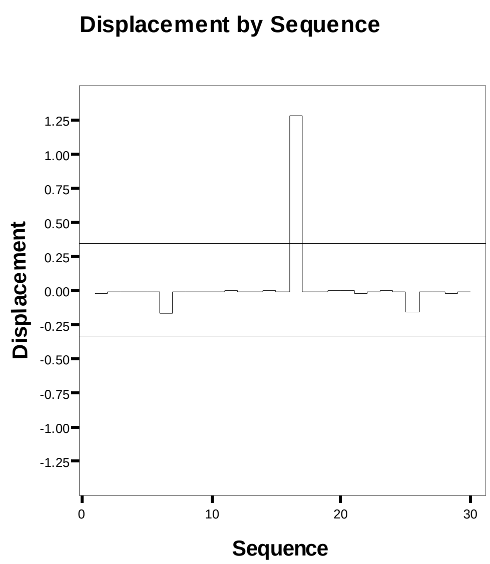
30 item test with 200 candidates
20 percent of the items drifted easy
for 20 percent of the candidates

Figure A – 3



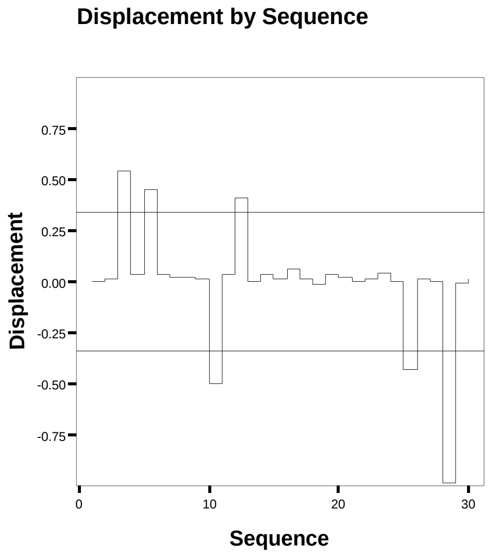
30 item test with 200 candidates
50 percent of the items drifted easy
for 50 percent of the candidates

Figure A – 4



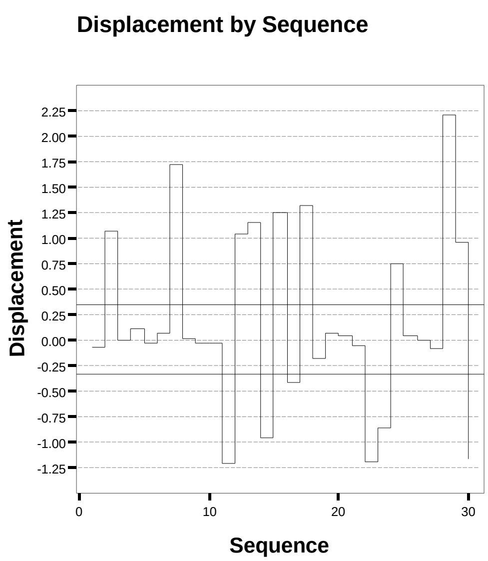
30 item test with 200 candidates
10 percent of the items drifted
half easy and half hard
for 10 percent of the candidates

Figure A-5



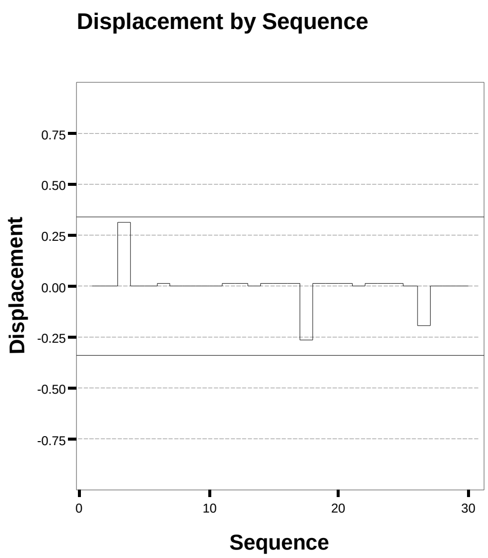
30 item test with 200 candidates
20 percent of the items drifted
half easy and half hard
for 20 percent of the candidates

Figure A-6



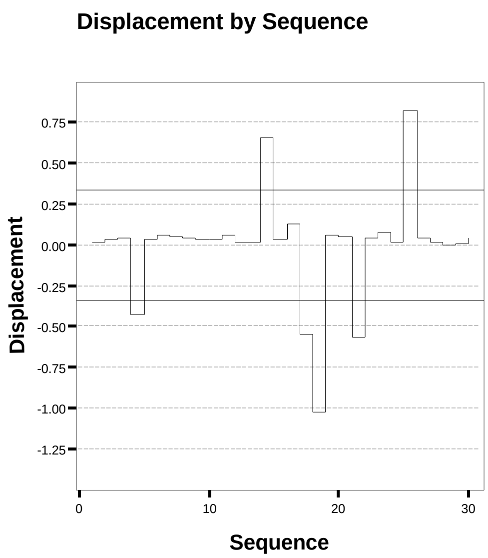
30 item test with 200 candidates
50 percent of the items drifted
half easy and half hard
for 50 percent of the candidates

Figure A-7



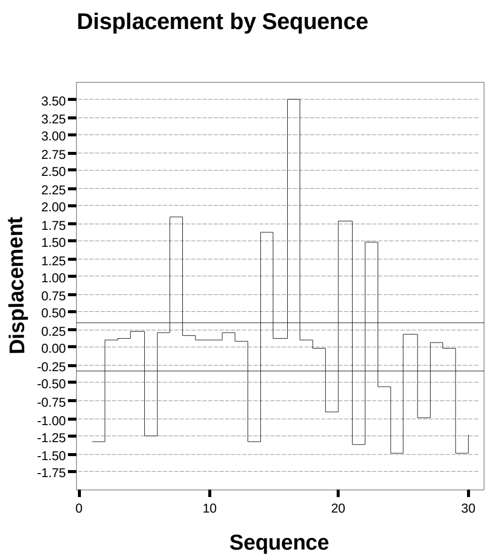
30 item test with 200 candidates
10 percent of the items drifted
30% harder and 70% easier
for 10 percent of the candidates

Figure A-8



30 item test with 200 candidates
20 percent of the items drifted
30% harder and 70% easier
for 20 percent of the candidates

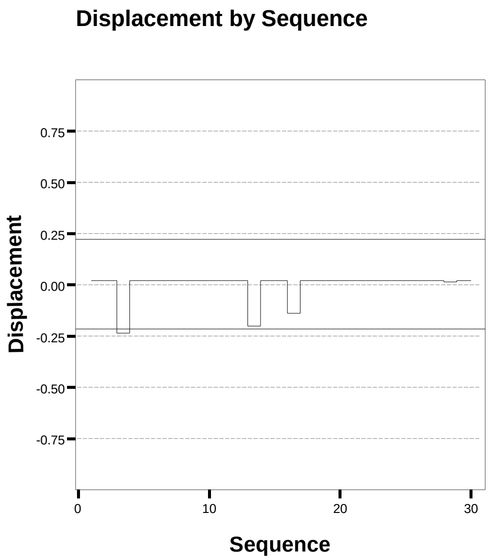
Figure A-9



30 item test with 200 candidates
50 percent of the items drifted
30% harder and 70% easier
for 50 percent of the candidates

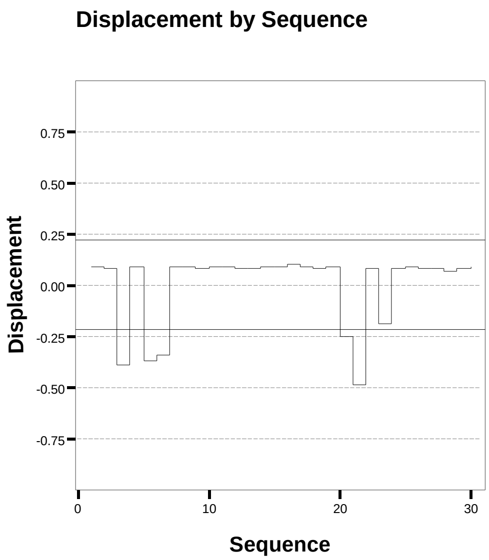
Appendix B: 30 items by 500 candidates

Figure B-1



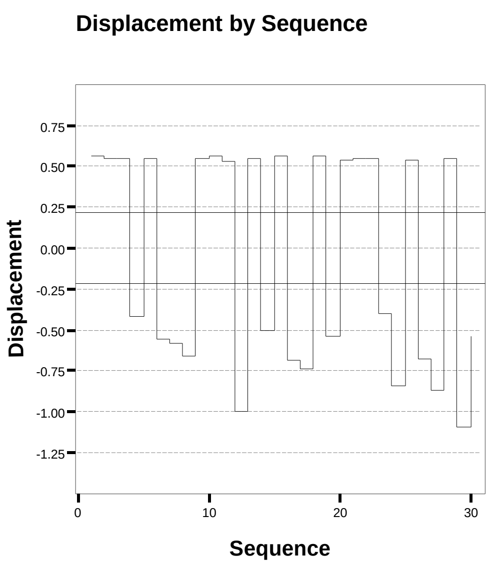
30 item test with 500 candidates
10 percent of the items drifted easy
for 10 percent of the candidates

Figure B-2



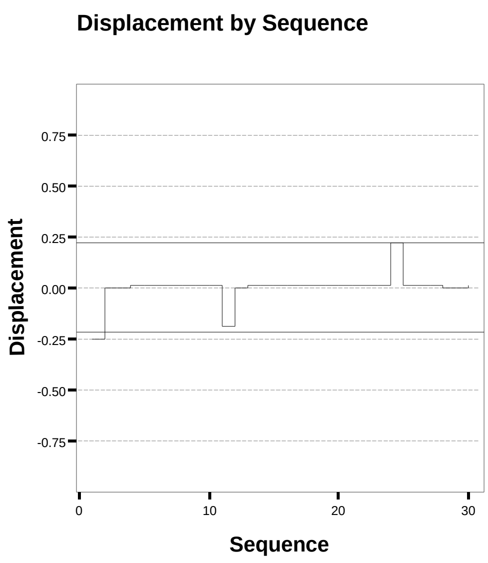
30 item test with 500 candidates
20 percent of the items drifted easy
for 20 percent of the candidates

Figure B-3



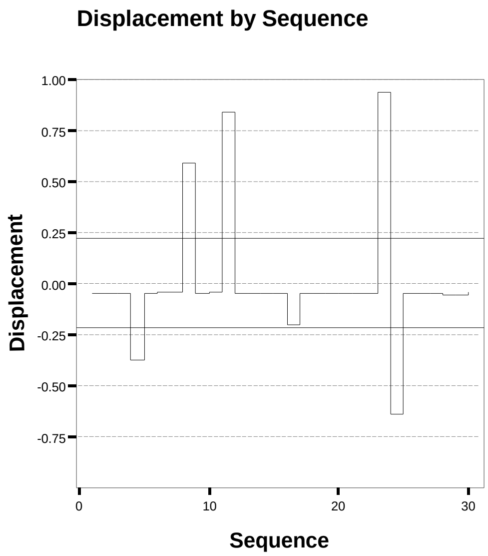
30 item test with 500 candidates
50 percent of the items drifted easy
for 50 percent of the candidates

Figure B-4



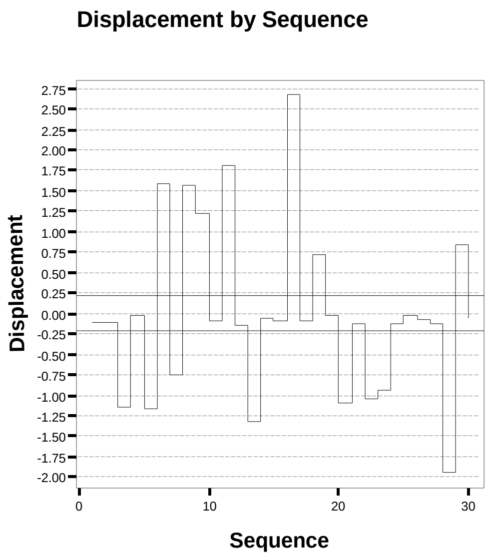
30 item test with 500 candidates
10 percent of the items drifted
half easy and half hard
for 10 percent of the candidates

Figure B-5



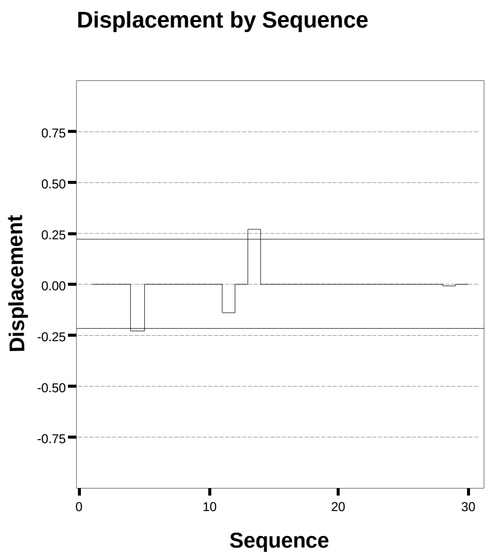
30 item test with 500 candidates
20 percent of the items drifted
half easy and half hard
for 20 percent of the candidates

Figure B-6



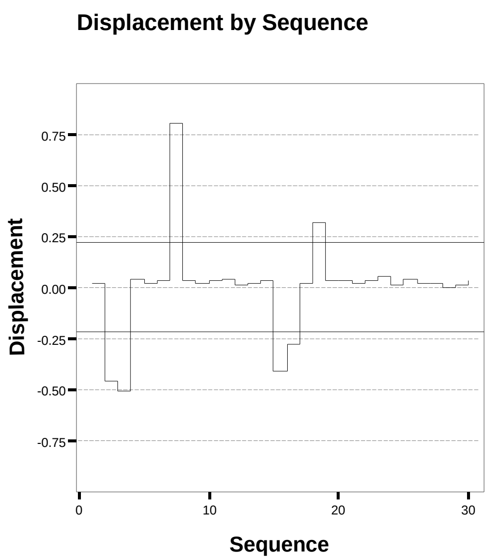
30 item test with 500 candidates
50 percent of the items drifted
half easy and half hard
for 50 percent of the candidates

Figure B-7



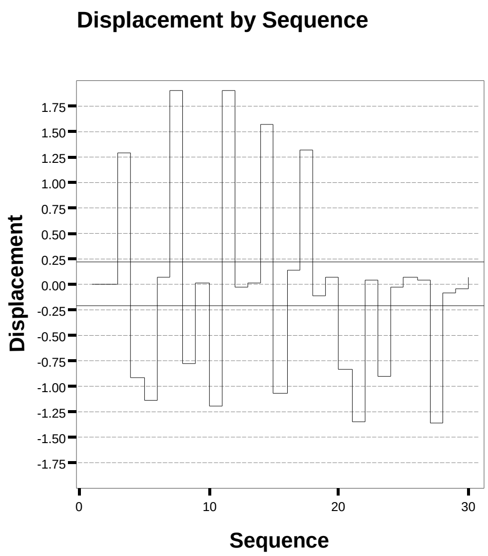
30 item test with 500 candidates
10 percent of the items drifted
30% harder and 70% easier
for 10 percent of the candidates

Figure B-8



30 item test with 500 candidates
20 percent of the items drifted
30% harder and 70% easier
for 20 percent of the candidates

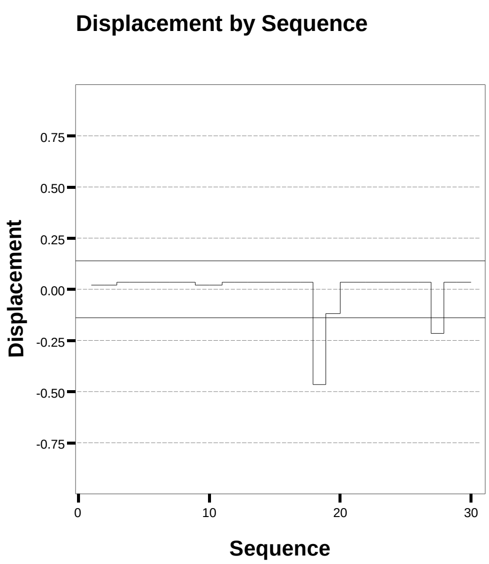
Figure B-9



30 item test with 500 candidates
50 percent of the items drifted
30% harder and 70% easier
for 50 percent of the candidates

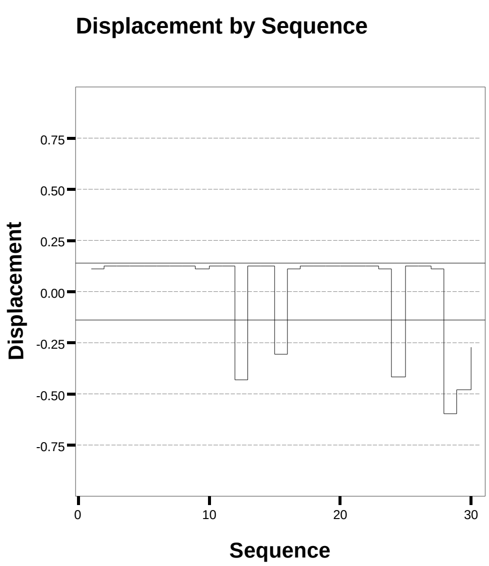
Appendix C: 30 items vs 1000 candidates

Figure C-1



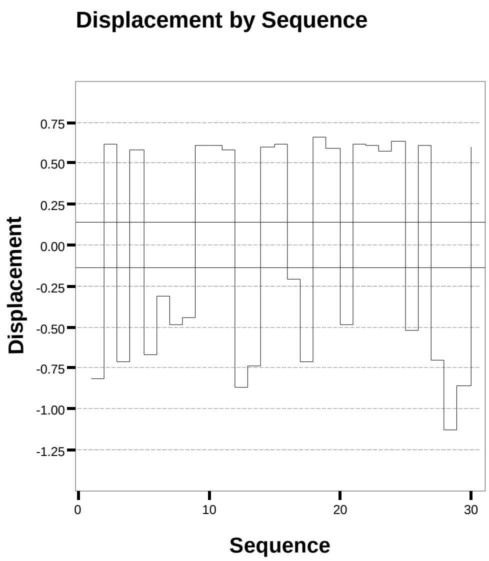
30 item test with 1000 candidates
10 percent of the items drifted easy
for 10 percent of the candidates

Figure C-2



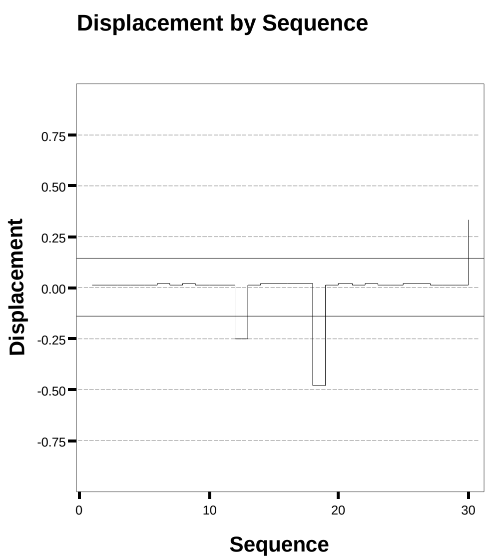
30 item test with 1000 candidates
20 percent of the items drifted easy
for 20 percent of the candidates

Figure C-3



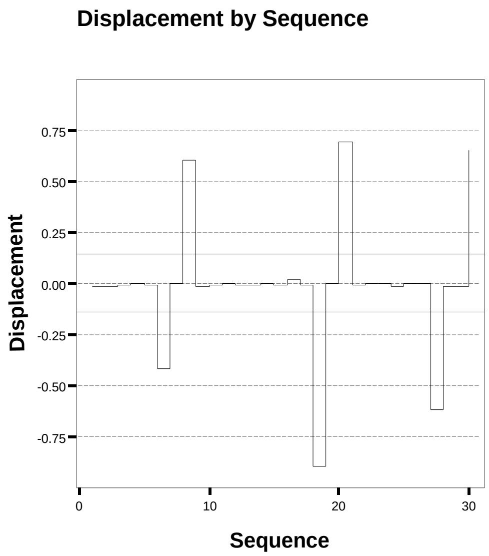
30 item test with 1000 candidates
50 percent of the items drifted easy
for 50 percent of the candidates

Figure C-4



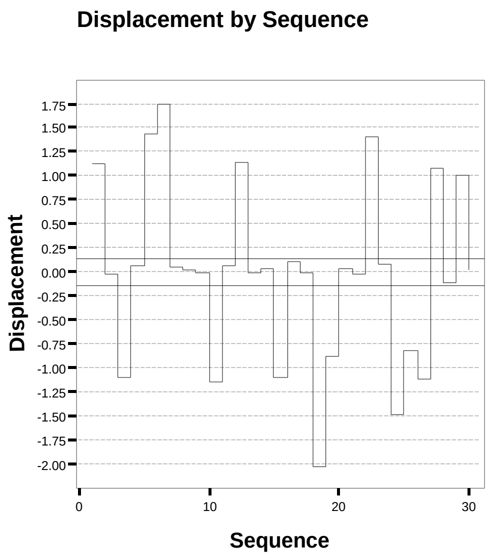
30 item test with 1000 candidates
10 percent of the items drifted
half harder and half easier
for 10 percent of the candidates

Figure C-5



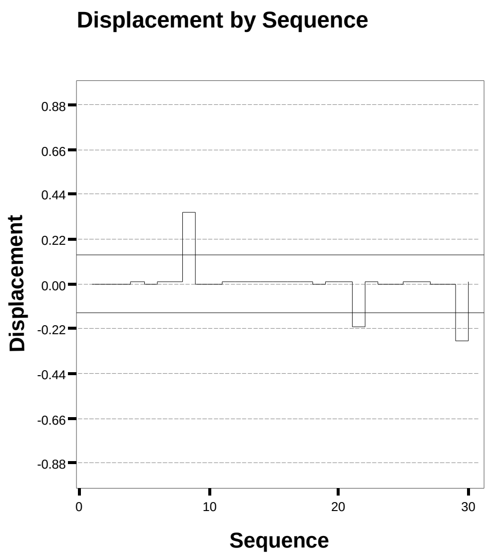
30 item test with 1000 candidates
20 percent of the items drifted
half harder and half easier
for 20 percent of the candidates

Figure C-6



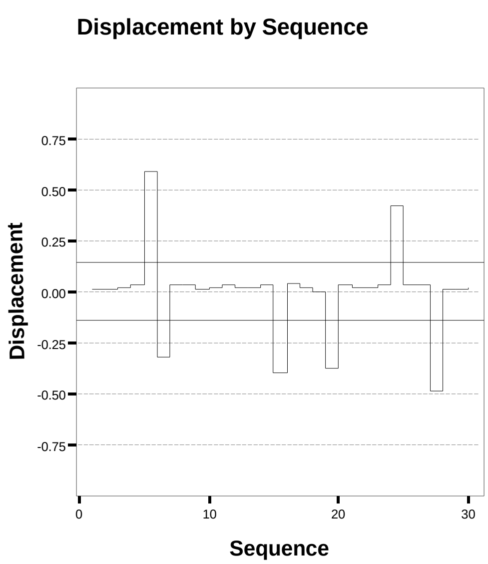
30 item test with 1000 candidates
50 percent of the items drifted
half harder and half easier
for 50 percent of the candidates

Figure C-7



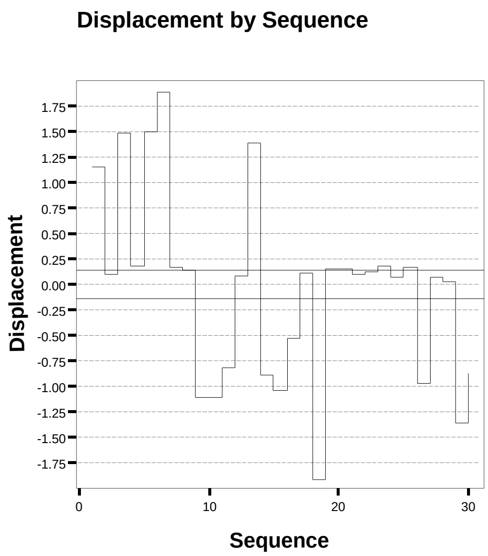
30 item test with 1000 candidates
10 percent of the items drifted
30% harder and 70% easier
for 10 percent of the candidates

Figure C-8



30 item test with 1000 candidates
20 percent of the items drifted
30% harder and 70% easier
for 20 percent of the candidates

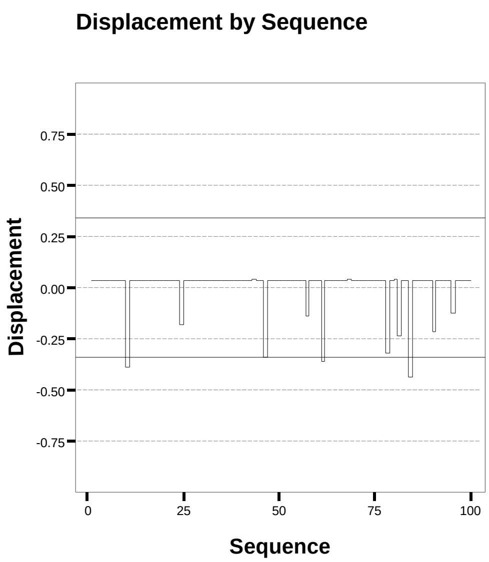
Figure C-9



30 item test with 1000 candidates
50 percent of the items drifted
30% harder and 70% easier
for 50 percent of the candidates

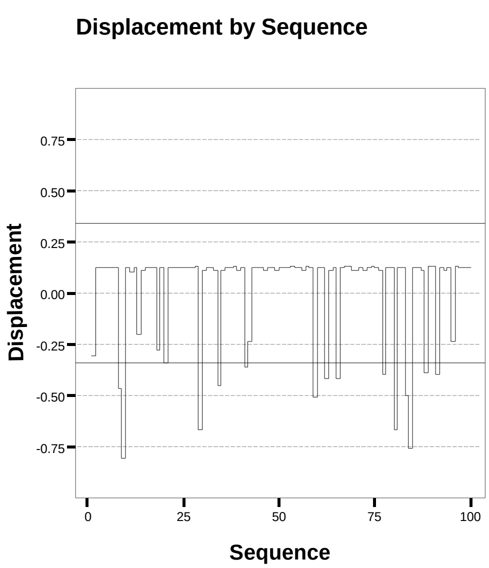
Appendix D; 100 items vs 200 candidates

Figure D-1



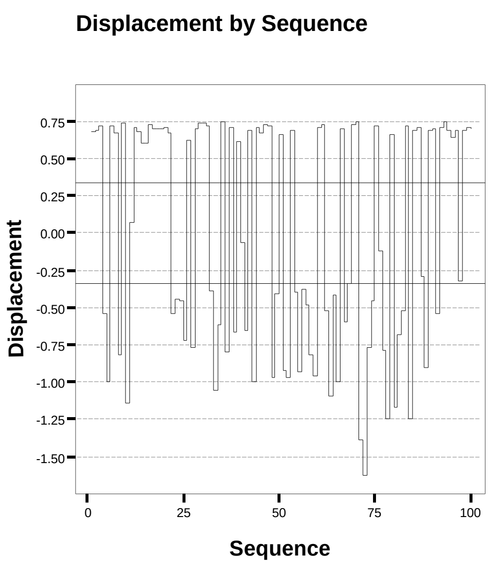
100 item test with 200 candidates
10 percent of the items drifted easier
for 10 percent of the candidates

Figure D-2



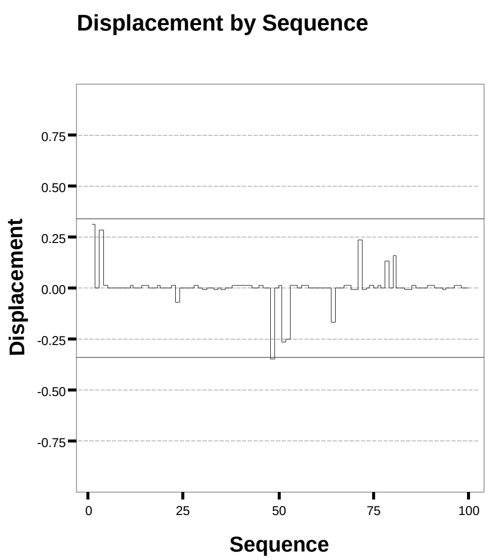
100 item test with 200 candidates
20 percent of the items drifted easier
for 20 percent of the candidates

Figure D-3



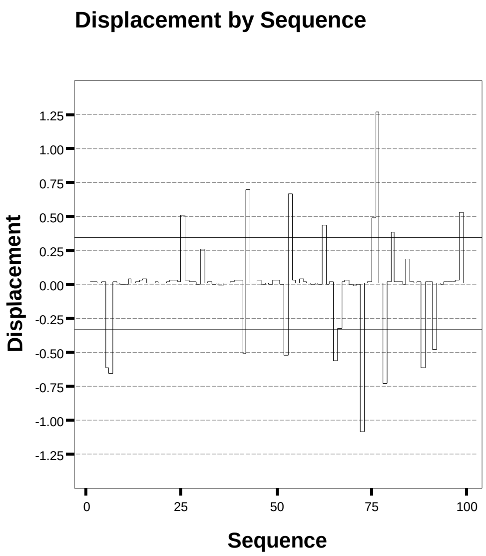
100 item test with 200 candidates
50 percent of the items drifted easier
for 50 percent of the candidates

Figure D-4



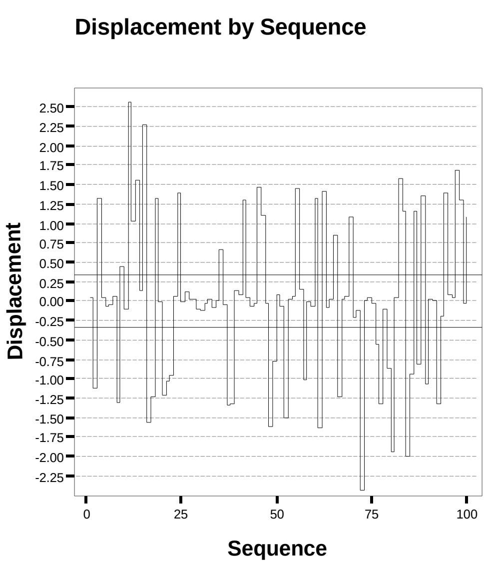
100 item test with 200 candidates
10 percent of the items drifted
half harder and half easier
for 10 percent of the candidates

Figure D-5



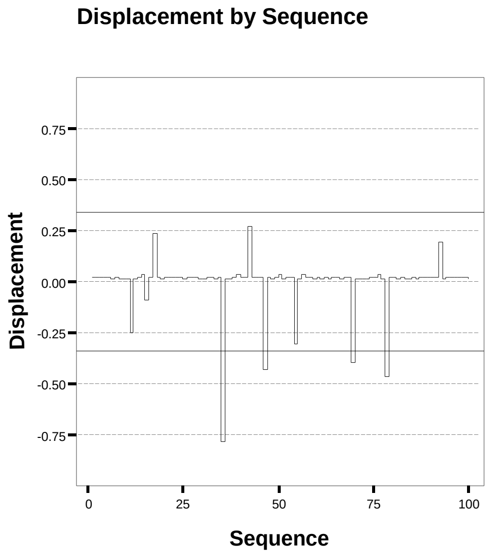
100 item test with 200 candidates
20 percent of the items drifted
half harder and half easier
for 20 percent of the candidates

Figure D-6



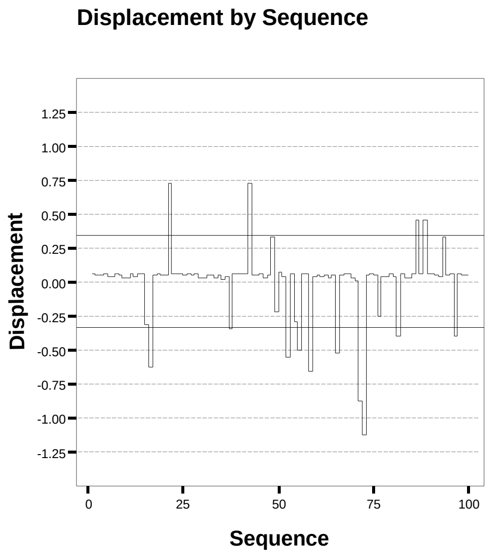
100 item test with 200 candidates
50 percent of the items drifted
half harder and half easier
for 50 percent of the candidates

Figure D-7



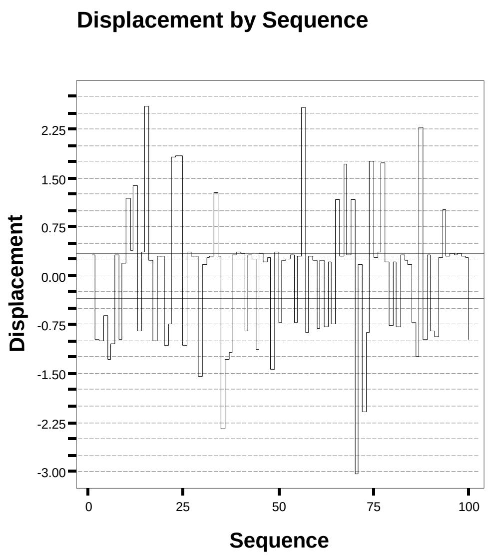
100 item test with 200 candidates
10 percent of the items drifted
30% harder and 70% easier
for 10 percent of the candidates

Figure D-8



100 item test with 200 candidates
20 percent of the items drifted
30% harder and 70% easier
for 20 percent of the candidates

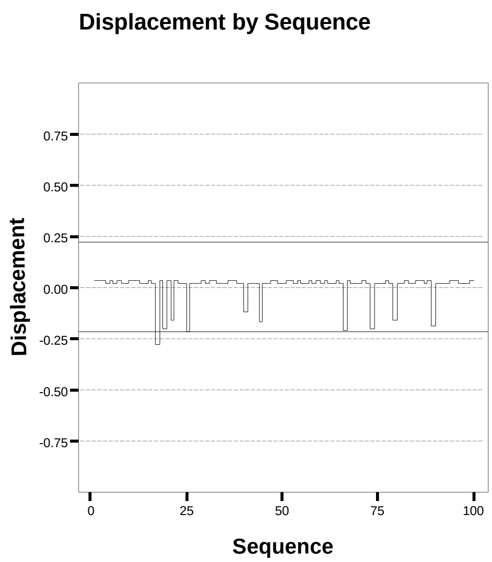
Figure D-9



100 item test with 200 candidates
50 percent of the items drifted
30% harder and 70% easier
for 50 percent of the candidates

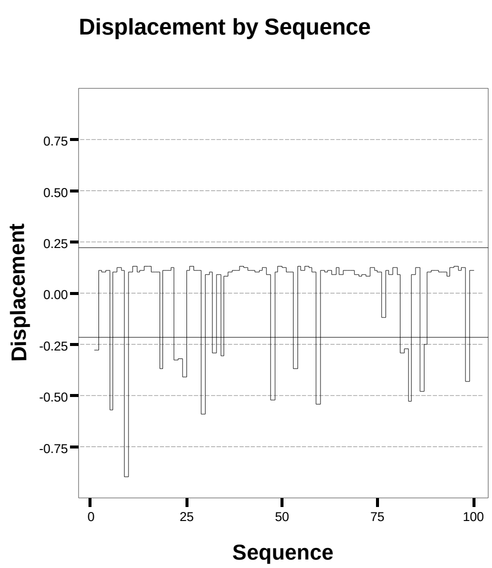
Appendix E: 100 items vs 500 candidates

Figure E-1



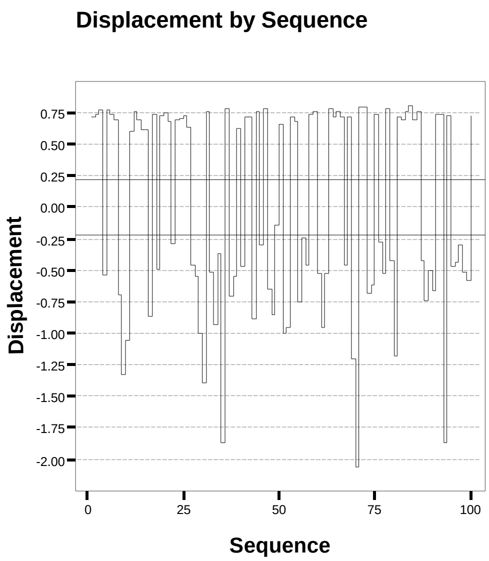
100 item test with 500 candidates
10 percent of the items drifted easier
for 10 percent of the candidates

Figure E-2



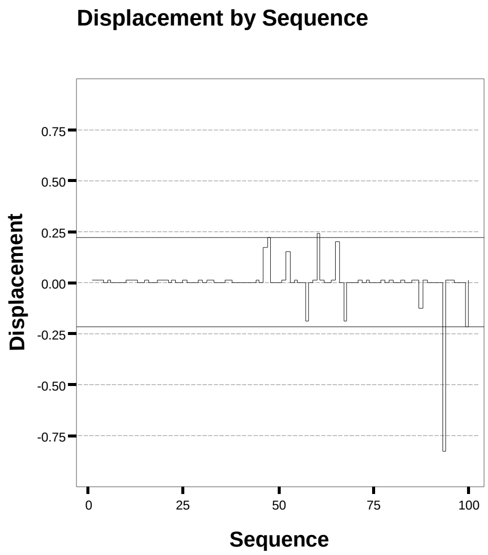
100 item test with 500 candidates
20 percent of the items drifted easier
for 20 percent of the candidates

Figure E-3



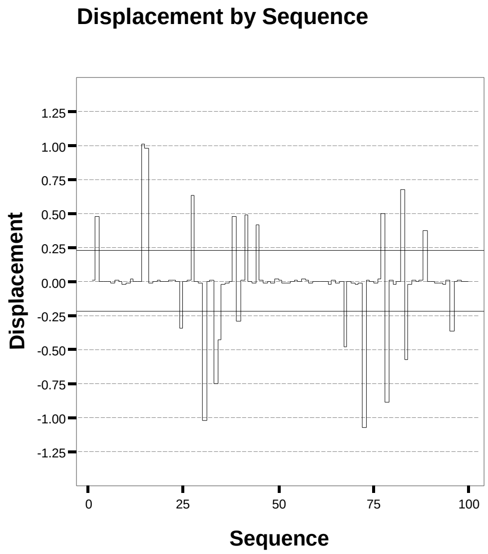
100 item test with 500 candidates
50 percent of the items drifted easier
for 50 percent of the candidates

Figure E-4



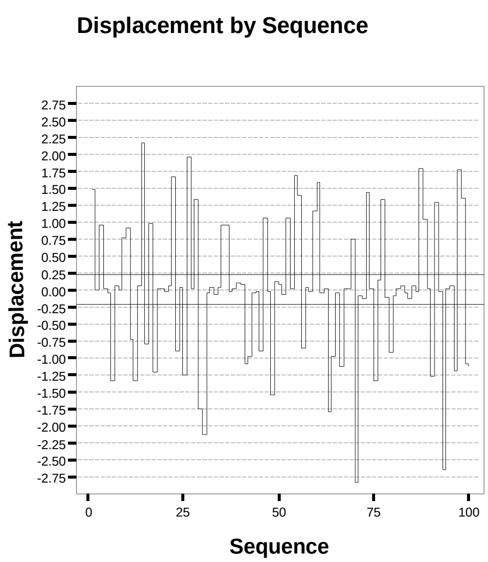
100 item test with 500 candidates
10 percent of the items drifted
half harder and half easier
for 10 percent of the candidates

Figure E-5



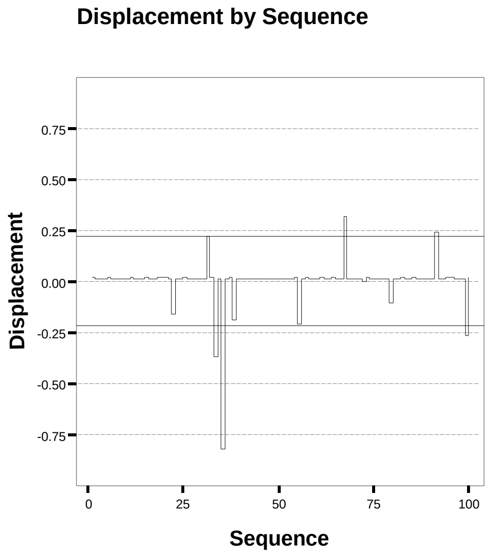
100 item test with 500 candidates
20 percent of the items drifted
half harder and half easier
for 20 percent of the candidates

Figure E-6



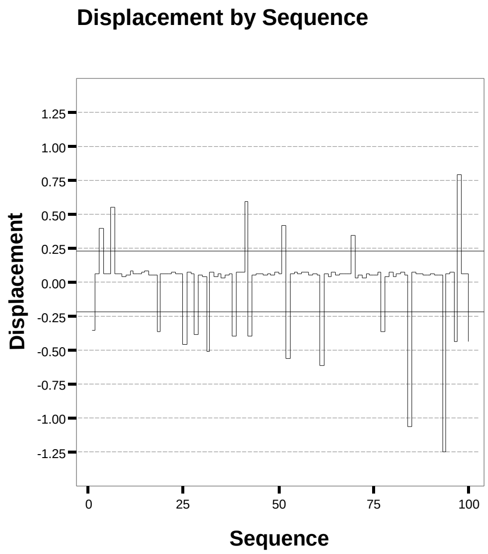
100 item test with 500 candidates
50 percent of the items drifted
half harder and half easier
for 50 percent of the candidates

Figure E-7



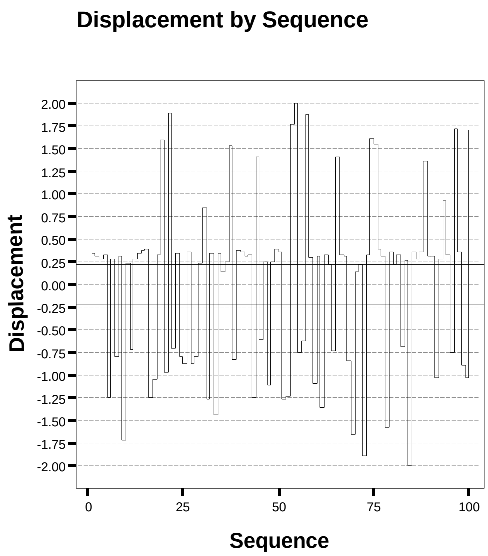
100 item test with 500 candidates
10 percent of the items drifted
30% harder and 70% easier
for 10 percent of the candidates

Figure E-8



100 item test with 500 candidates
20 percent of the items drifted
30% harder and 70% easier
for 20 percent of the candidates

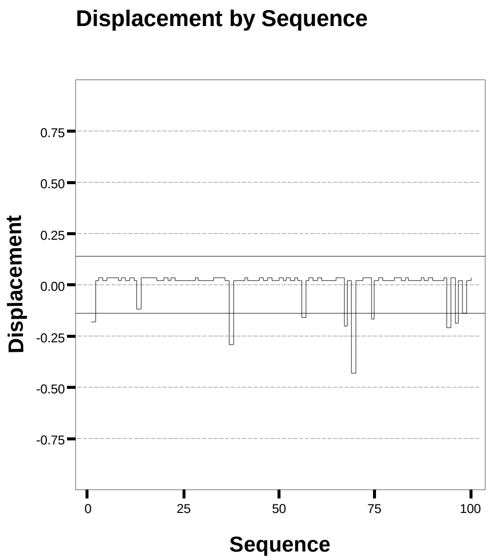
Figure E-9



100 item test with 500 candidates
50 percent of the items drifted
30% harder and 70% easier
for 50 percent of the candidates

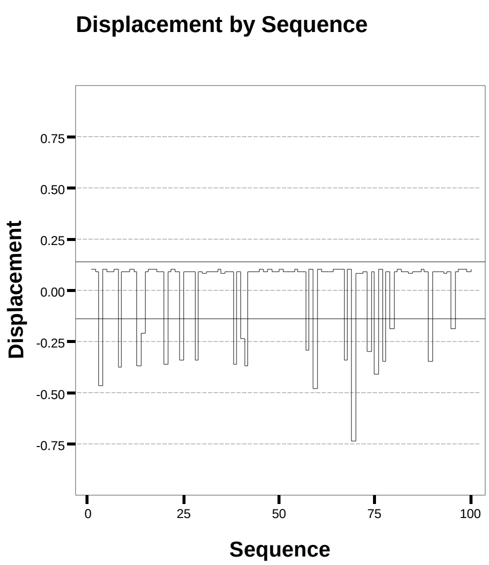
Appendix F: 100 items vs 1000 candidates

Figure F-1



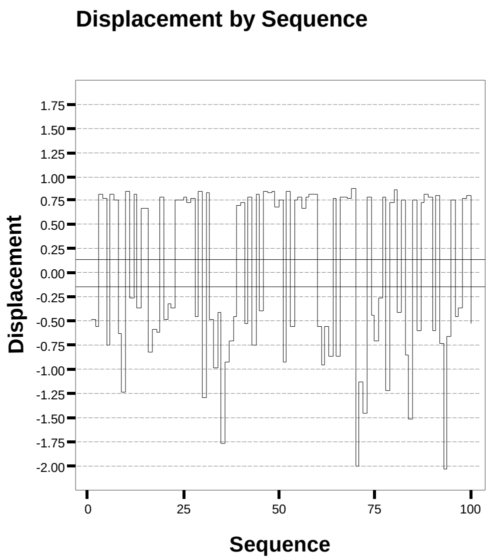
100 item test with 1000 candidates
10 percent of the items drifted easier
for 10 percent of the candidates

Figure F-2



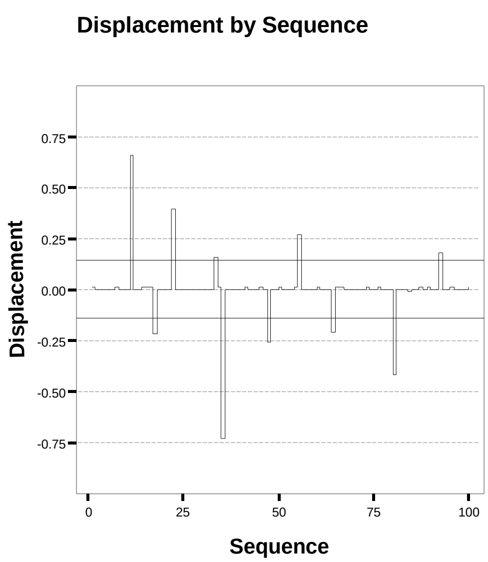
100 item test with 1000 candidates
20 percent of the items drifted easier
for 20 percent of the candidates

Figure F-3



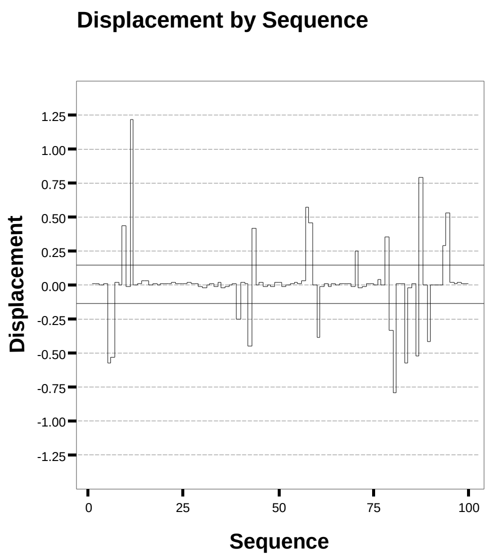
100 item test with 1000 candidates
50 percent of the items drifted easier
for 50 percent of the candidates

Figure F-4



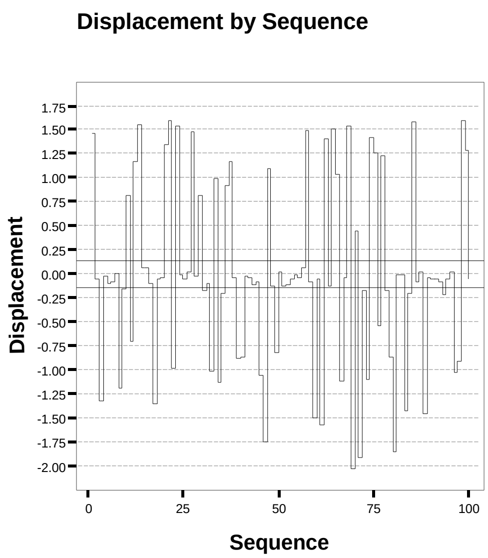
100 item test with 1000 candidates
10 percent of the items drifted
half harder and half easier
for 10 percent of the candidates

Figure F-5



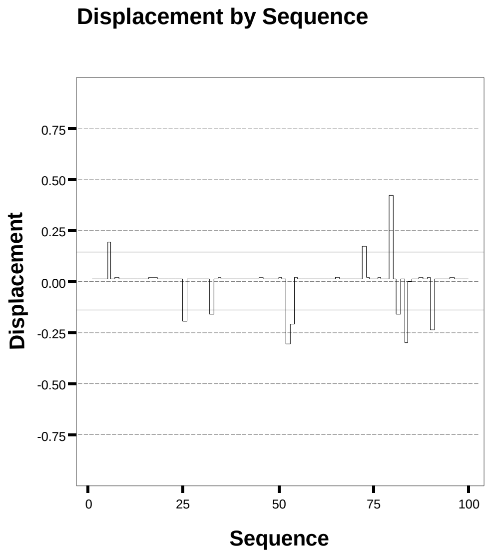
100 item test with 1000 candidates
20 percent of the items drifted
half harder and half easier
for 20 percent of the candidates

Figure F-6



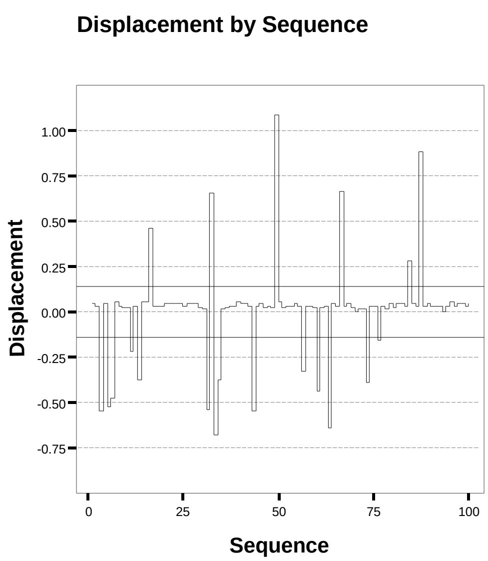
100 item test with 1000 candidates
50 percent of the items drifted
half harder and half easier
for 50 percent of the candidates

Figure F-7



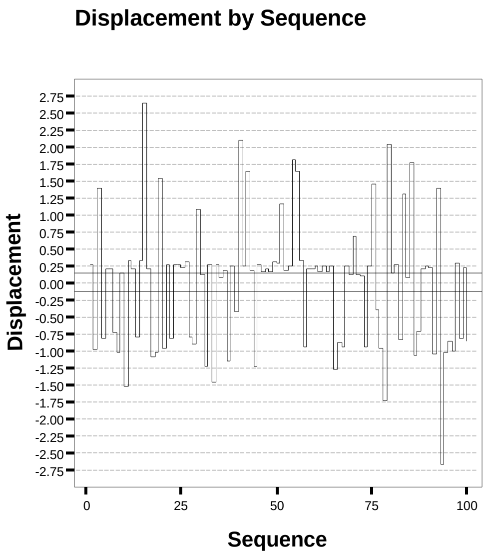
100 item test with 1000 candidates
10 percent of the items drifted
30% harder and 70% easier
for 10 percent of the candidates

Figure F-8



100 item test with 1000 candidates
20 percent of the items drifted
30% harder and 70% easier
for 20 percent of the candidates

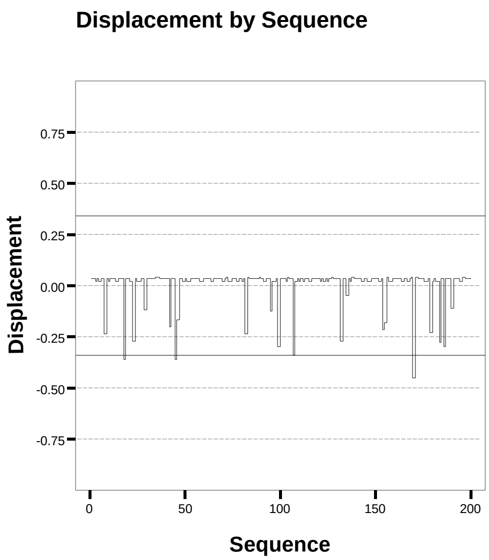
Figure F-9



100 item test with 1000 candidates
50 percent of the items drifted
30% harder and 70% easier
for 50 percent of the candidates

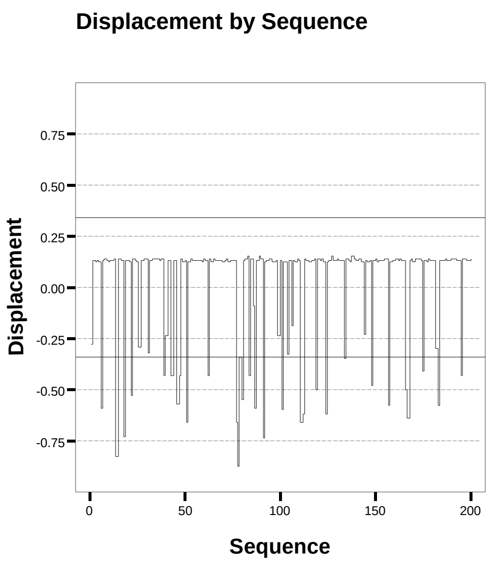
Appendix G: 200 items vs 200 candidates

Figure G-1



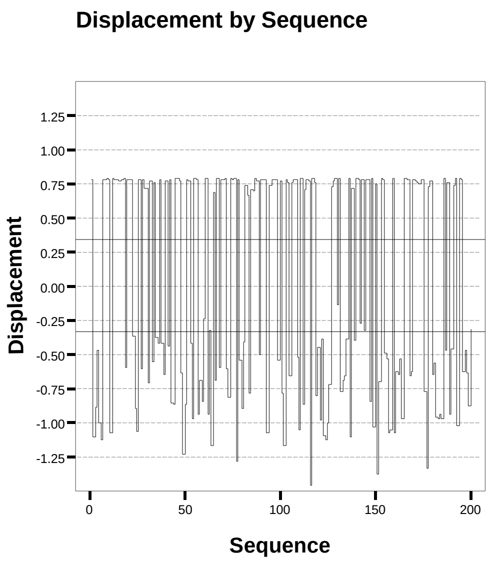
200 item test with 200 candidates
10 percent of the items drifted easier
for 10 percent of the candidates

Figure G-2



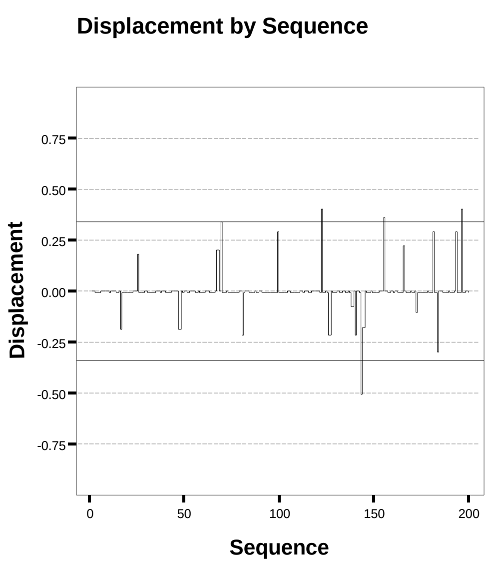
200 item test with 200 candidates
20 percent of the items drifted easier
for 20 percent of the candidates

Figure G-3



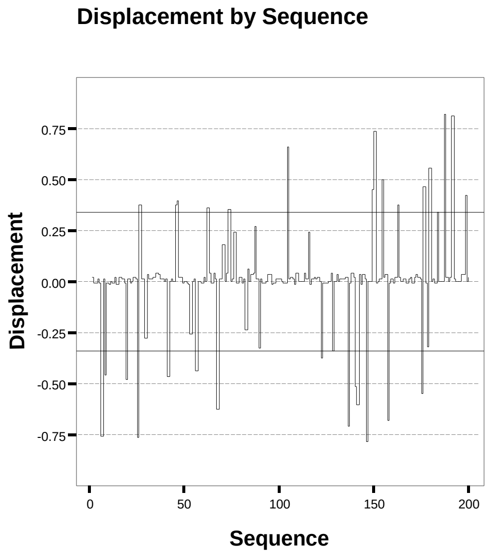
200 item test with 200 candidates
50 percent of the items drifted easier
for 50 percent of the candidates

Figure G-4



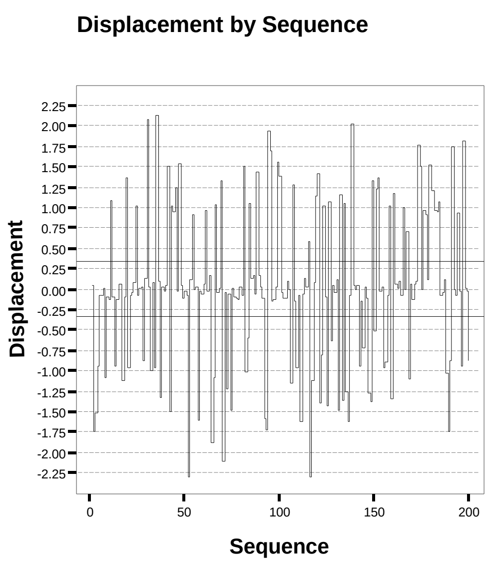
200 item test with 200 candidates
10 percent of the items drifted
half harder and half easier
for 10 percent of the candidates

Figure G-5



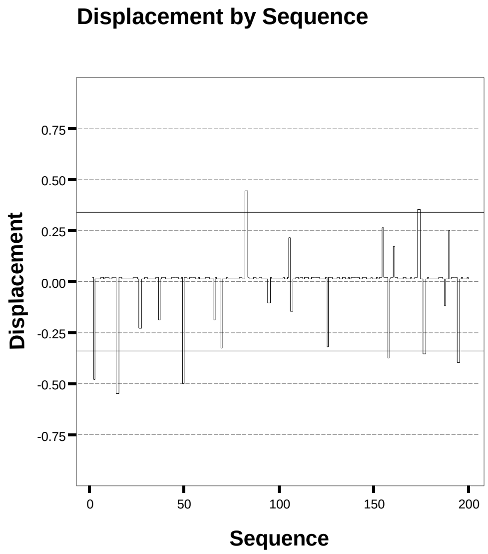
200 item test with 200 candidates
20 percent of the items drifted
half harder and half easier
for 20 percent of the candidates

Figure G-6



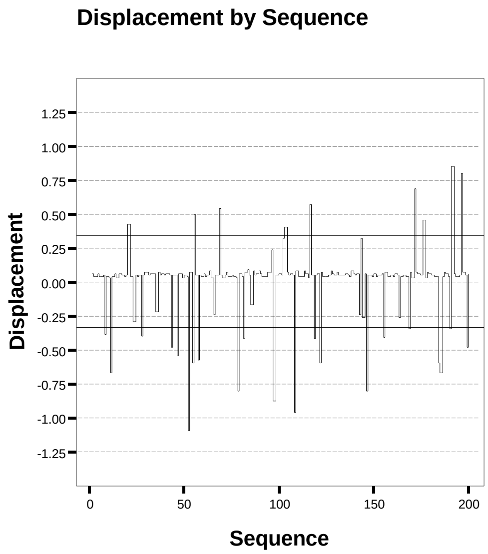
200 item test with 200 candidates
50 percent of the items drifted
half harder and half easier
for 50 percent of the candidates

Figure G-7



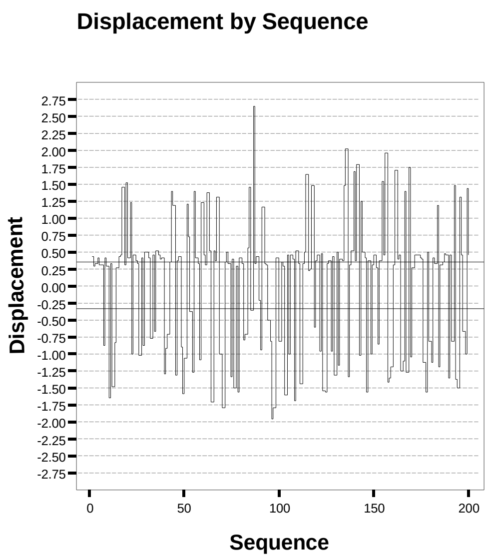
200 item test with 200 candidates
10 percent of the items drifted
30% harder and 70% easier
for 10 percent of the candidates

Figure G-8



200 item test with 200 candidates
20 percent of the items drifted
30% harder and 70% easier
for 20 percent of the candidates

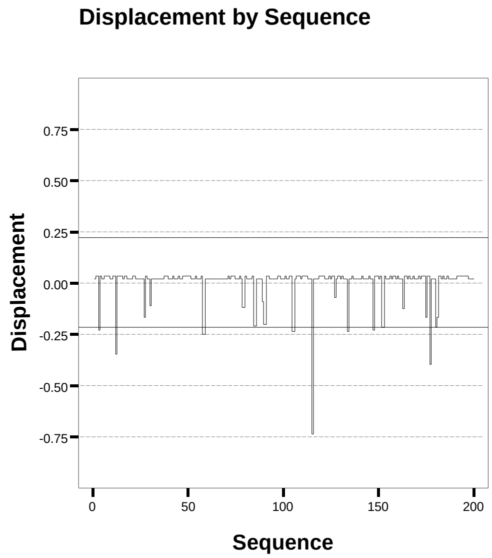
Figure G-9



200 item test with 200 candidates
50 percent of the items drifted
30% harder and 70% easier
for 50 percent of the candidates

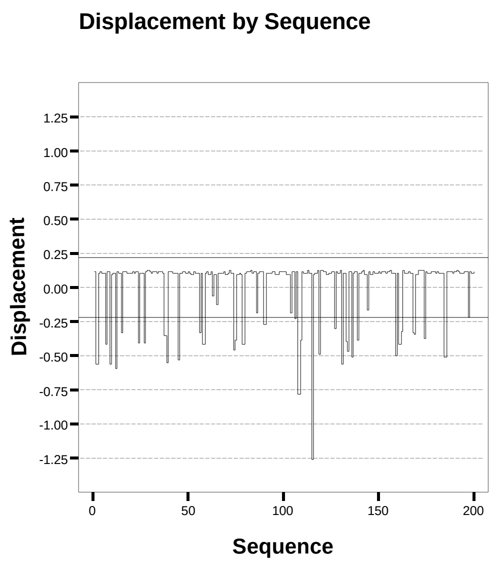
Appendix H: 200 items vs. 500 candidates

Figure H-1



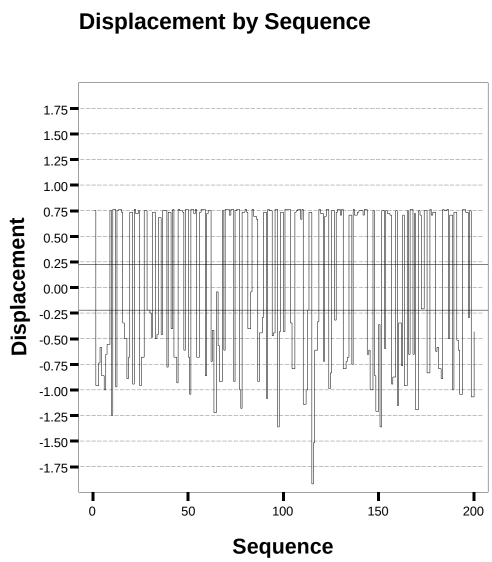
200 item test with 500 candidates
10 percent of the items drifted easier
for 10 percent of the candidates

Figure H-2



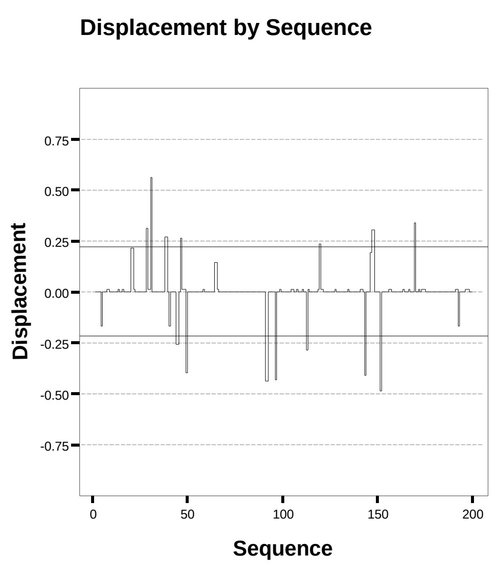
200 item test with 500 candidates
20 percent of the items drifted easier
for 20 percent of the candidates

Figure H-3



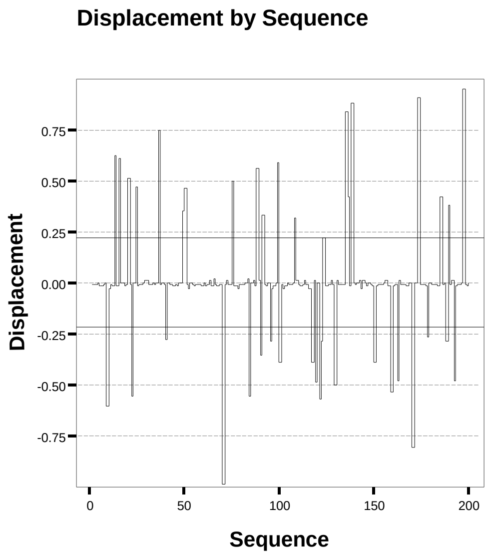
200 item test with 500 candidates
50 percent of the items drifted easier
for 50 percent of the candidates

Figure H-4



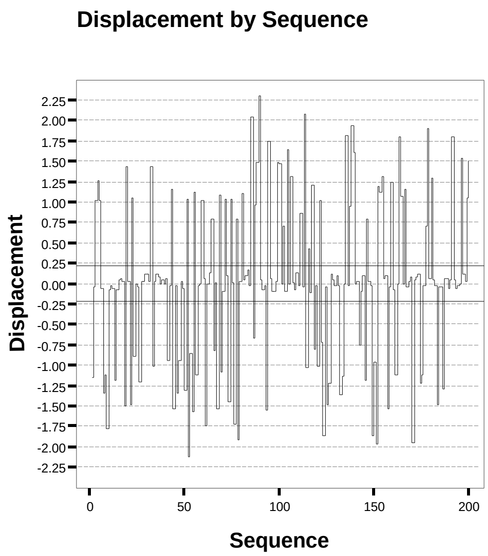
200 item test with 500 candidates
10 percent of the items drifted
half harder and half easier
for 10 percent of the candidates

Figure H-5



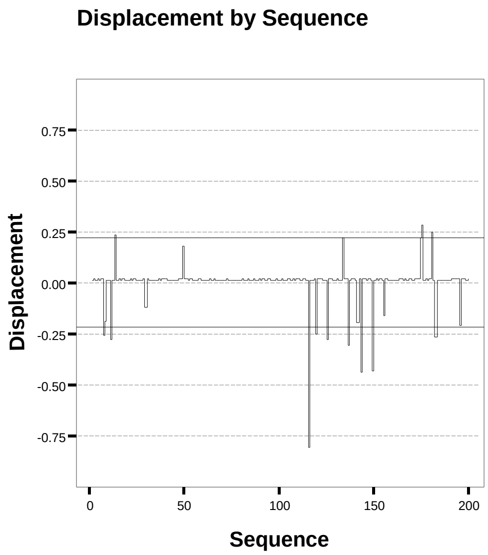
200 item test with 500 candidates
20 percent of the items drifted
half harder and half easier
for 20 percent of the candidates

Figure H-6



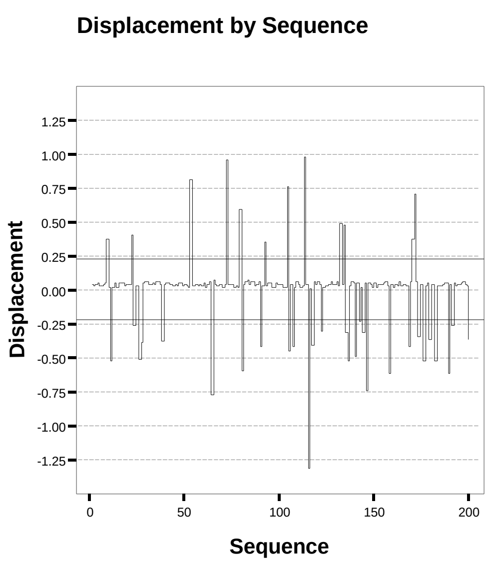
200 item test with 500 candidates
50 percent of the items drifted
half harder and half easier
for 50 percent of the candidates

Figure H-7



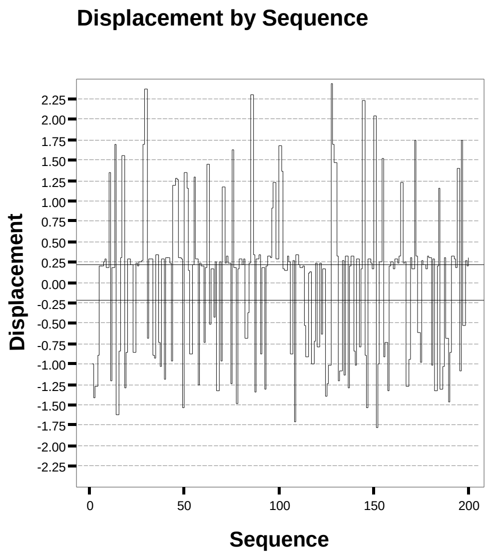
200 item test with 500 candidates
10 percent of the items drifted
30% harder and 70% easier
for 10 percent of the candidates

Figure H-8



200 item test with 500 candidates
20 percent of the items drifted
30% harder and 70% easier
for 20 percent of the candidates

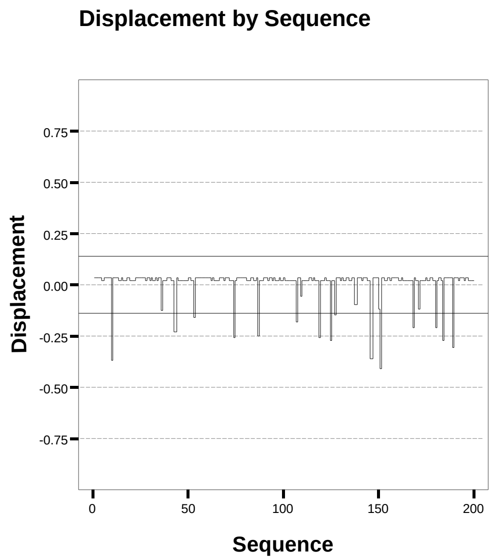
Figure H-9



200 item test with 500 candidates
50 percent of the items drifted
30% harder and 70% easier
for 50 percent of the candidates

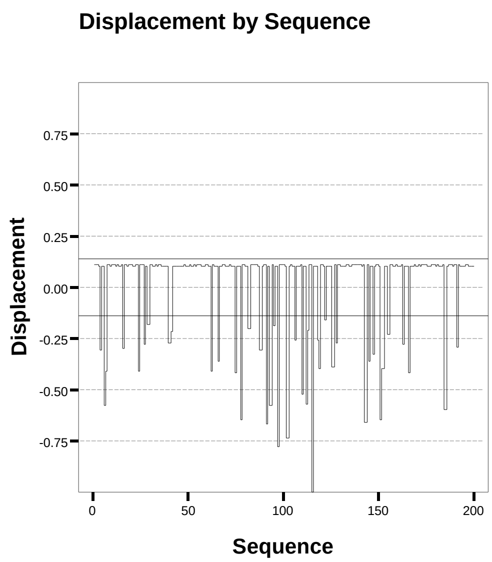
Appendix I: 200 items vs1000 candidates

Figure I-1



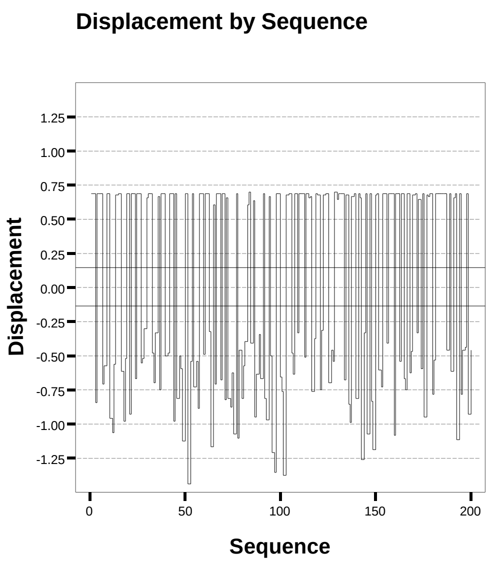
200 item test with 1000 candidates
10 percent of the items drifted easier
for 10 percent of the candidates

Figure I-2



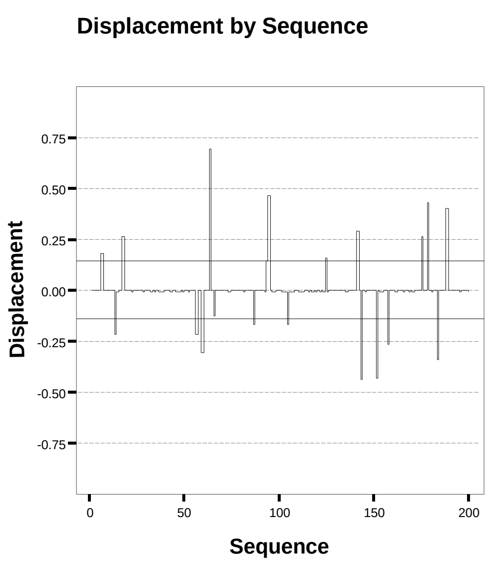
200 item test with 1000 candidates
20 percent of the items drifted easier
for 20 percent of the candidates

Figure I-3



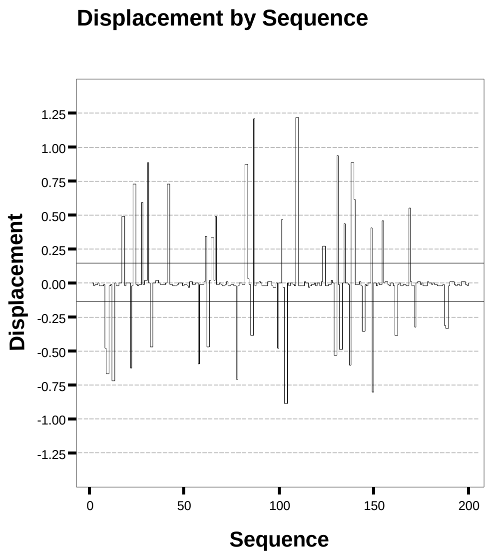
200 item test with 1000 candidates
50 percent of the items drifted easier
for 50 percent of the candidates

Figure I-4



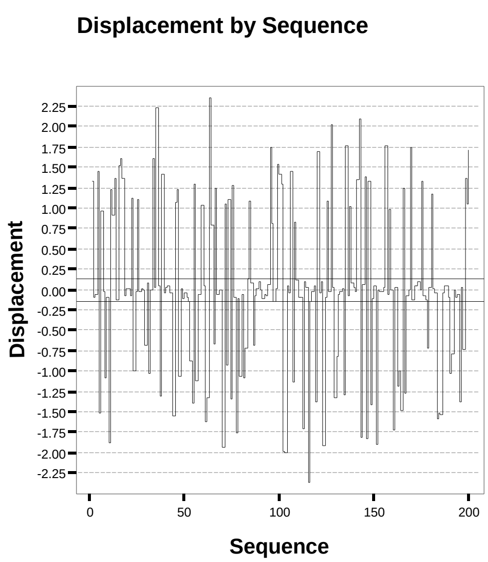
200 item test with 1000 candidates
10 percent of the items drifted
half harder and half easier
for 10 percent of the candidates

Figure I-5



200 item test with 1000 candidates
20 percent of the items drifted
half harder and half easier
for 20 percent of the candidates

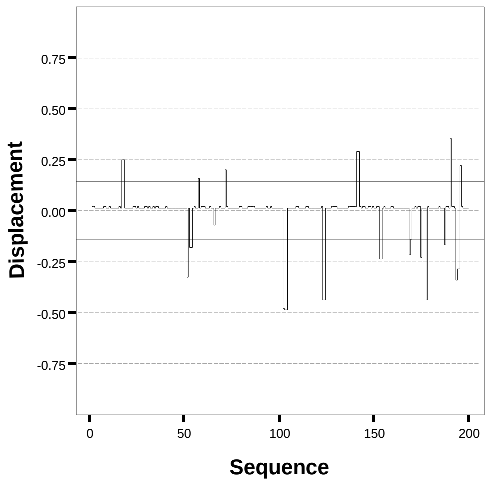
Figure I-6



200 item test with 1000 candidates
50 percent of the items drifted
half harder and half easier
for 50 percent of the candidates

Figure I-7

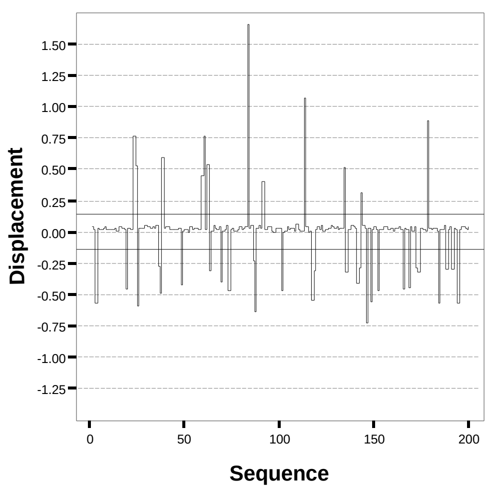
Displacement by Sequence



200 item test with 1000 candidates
10 percent of the items drifted
30% harder and 70% easier
for 10 percent of the candidates

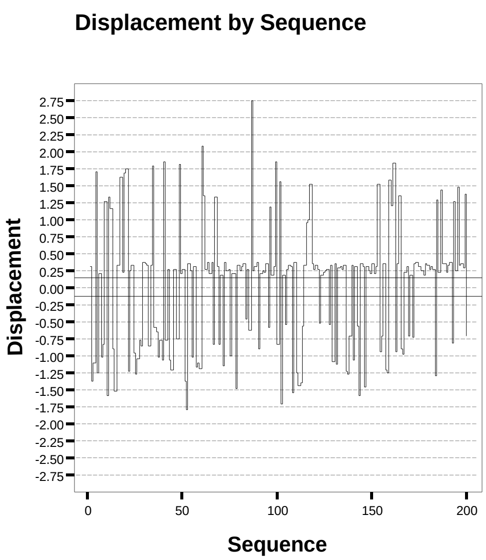
Figure I-8

Displacement by Sequence



200 item test with 1000 candidates
20 percent of the items drifted
30% harder and 70% easier
for 20 percent of the candidates

Figure I-9



200 item test with 1000 candidates
50 percent of the items drifted
30% harder and 70% easier
for 50 percent of the candidates