

Third-Party Checking of 2013 Scaling and Equating for the Kentucky Performance Rating for Educational Progress (K-PREP) Tests

Prepared for: Kentucky Department of Education
Office of Assessment and Accountability
17th Floor, Capital Plaza Tower
500 Mero Street
Frankfort, KY 40601

Prepared by: Human Resources Research Organization
10503 Timberwood Circle, Suite 101
Louisville, KY 40223

Authors: Bethany H. Bynum
Erin Smith
Arthur A. Thacker

Date: November 4, 2013



Human Resources Research Organization (HumRRO)
66 Canal Center Plaza, Suite 700, Alexandria, Virginia 22314-1591
Phone: 703.549.3611 | Fax: 703.549.9661 | www.humrro.org

Third-Party Checking of 2013 Scaling and Equating for the Kentucky Performance Rating for Educational Progress (K-PREP) Tests

Prepared for: Kentucky Department of Education
Office of Assessment and Accountability
17th Floor, Capital Plaza Tower
500 Mero Street
Frankfort, KY 40601

Prepared by: Human Resources Research Organization
10503 Timberwood Circle, Suite 101
Louisville, KY 40223

Authors: Bethany H. Bynum
Erin Smith
Arthur A. Thacker

Date: November 4, 2013



THIRD-PARTY CHECKING OF 2013 SCALING AND EQUATING FOR THE KENTUCKY PERFORMANCE RATING FOR EDUCATIONAL PROGRESS (K-PREP) TESTS

Table of Contents

Introduction	1
Sample Identification and File Construction.....	1
Calibration and Scaling Procedures.....	2
Equating Procedures	2
Raw-score-to-Scale-Score Procedures	3
Verification of 2013 Scoring Tables	3
Documentation.....	3
Issues.....	4
Conclusion	54
References.....	6
Appendix A – Control File (Reading Grade 3)	A-1
Appendix B – Winsteps Item Parameter Files (Reading Grade 3).....	B-1
Appendix C – Robust Z Item Stability Analysis.....	C-1
Appendix D – Winsteps Anchor File (Grade 3 Reading).....	D-1
Appendix E – Winsteps Score File (Grade 3 Reading).....	E-1
Appendix F – Comparison of Files Output (Reading Grade 3)	F-1

Executive Summary

Pearson and HumRRO independently calibrated, scaled and equated the 2013 Kentucky Performance Rating for Educational Progress (K-PREP) assessments and produced the raw-score-to-theta-score tables to be applied to students' test results. Results calculated by HumRRO were identical to those calculated by Pearson (M. Johnson, email communication, August 5, 2013). Given that HumRRO's results were identical to those of Pearson, we are assured that Pearson did not commit processing errors.

THIRD-PARTY CHECKING OF 2013 SCALING AND EQUATING FOR THE KENTUCKY PERFORMANCE RATING FOR EDUCATIONAL PROGRESS (K-PREP) TESTS

Introduction

In 2012, Kentucky transitioned from the Kentucky Core Content Test (KCCT) to the K-PREP system for spring testing. This transition represented a significant departure from the prior assessment system. The 3 parameter logistic IRT model was replaced with a Rasch model, a new item-type (i.e., short-constructed-response) was added to the assessments, a new scale-score reporting system was developed for sub-scores, and new cut scores were identified for the reading and mathematics assessments. The transition was also accompanied by a new primary testing contractor, Pearson. As a result, HumRRO's third-party checking process underwent significant changes to accommodate the transition¹.

Now in the second year of administration, equating was added to the process to permit comparison of the results across test years. The 2013 tests were equated to the 2012 tests using linking items. In this manner, comparable scores were produced for the 2013 K-PREP.

This report describes how student test responses for the 2013 K-PREP assessments were used to create scale scores and place students in Novice, Apprentice, Proficient or Distinguished (NAPD) performance categories. The complex analyses to accomplish these tasks were conducted independently, but cooperatively, by both HumRRO and Pearson staff members. Several interim checks were conducted during the analyses and any discrepancies between the two companies was investigated and ultimately resolved. This process was conducted transparently among Pearson, HumRRO, KDE, and Kentucky's psychometric consultant (Dr. Bill Auty of EdMeasure) via frequent email communications and daily conference calls. The process was guided by a specifications document created by Pearson² and regularly updated based on decisions before and during calibration. This documentation is vital for ensuring consistency of processing across years and will be a guiding document for subsequent years.

Sample Identification and File Construction

Kentucky selects most of its student population for use in the calibration sample for scaling and equating. However, some students are purposefully exempted. KDE established a set of invalidation codes for excluding students in the calibration file. Kentucky's exemption rules only apply to students who receive accommodations (e.g., Braille forms, audio, large print, etc.) and students with duplicate records (the same identification number and name). The accommodated students receive scores, but are simply omitted from the calibration sample. Pearson and HumRRO verified n-counts after this step.

The next step was to format all of the grade/subject files to be read into the Winsteps IRT program and create Winsteps control files to read the student responses and estimate parameters. A sample control file is provided in Appendix A. HumRRO created specialized SAS programs to generate all input and control files automatically. The item documentation file was used to specify item types, location, keys, item use (field test vs. operational items), and other important information. HumRRO and Pearson did not share programming or methodology for

¹ For additional details on how the assessment system and third-party checking procedures changed, see Bynum and Thacker (2013).

² K-PREP Sp13 CES Specs_KDE v1.5.doc.

creating the input and control files. Both companies did use the same student data files (containing all student responses). HumRRO followed the guidance provided by Pearson (with input from KDE) regarding the treatment of blank responses, condition codes, etc. in creating the input data files.

Calibration and Scaling Procedures

Once input and control files were prepared, Winsteps was used to calibrate items. Multiple-choice items were fit to the Rasch measurement model and constructed-response items (short constructed response and extended response items) were fit to the Partial Credit Model (PCM). Both types of items were simultaneously calibrated in Winsteps and item difficulty parameters (logits) were produced. “Step parameters” were also produced for constructed response items. Step parameters tell us how the various points possible on the item relate to the item’s overall difficulty and are important for generating scoring tables. These parameters are produced on the theta scale (a commonly used scale with a mean of 0 and a standard deviation of 1). Appendix B contains an example of item parameters for one grade subject (logits and step parameters).

Equating Procedures

Two types of equating occurred for the K-PREP: (a) forms equating within a given test administration year and (b) equating across test administration years using common anchor items. The first of these, forms equating, is accomplished by calibrating all of the items for a given grade/subject together. By calibrating all of the items together (i.e., across all forms), this effectively equates the various forms for a given grade/subject such that test scores on form 2 and form 3, for example, are interchangeable in terms of difficulty. In other words, a student should get about the same score regardless of which form he/she takes.

In addition to the need to equate the forms of a test within a given year, there is also the need for the current year’s scores to be comparable to scores from prior years. For 2013, we equated to the 2012 scale for Reading, Mathematics, Science and Social Studies. Kentucky uses a common-item anchor design to equate K-PREP scores across years. The anchor items are “internal” in the sense that they are dispersed across forms rather than externally located in a separate anchor item form. Both multiple-choice and open-response items are designated as anchor items for equating for all grades and subjects.

Equating across test administrations involved three steps. First, an initial calibration was run with all items freely estimated. Second, an item stability check was performed on the linking item parameters using the Robust Z statistical procedure (Huynh, 2000; Huynh & Rawls, 2009; Huynh & Meyer, 2010). The procedure compared the item difficulty estimates of the 2012 and 2013 linking items (*b*-parameter estimates for multiple-choice items and step parameters for open-ended items) to identify items which should be dropped. See Appendix C for an outline of the steps in the procedure and the criteria for exclusion. Any items flagged for removal were discussed with Pearson at this stage. Lastly, a final calibration was run using the 2012 item parameter estimates as anchor values for the linking items, thereby placing the 2013 tests onto the same measurement scale as the 2012 tests. The multiple-choice items were anchored to the 2012 item Rasch difficulty parameter estimates and the open-ended response items were anchored to both the 2012 item difficulty estimates and the category threshold measures that were computed for each step. These final (equated) item parameters were compared to Pearson’s parameter estimates.

Raw-score-to-Scale-Score Procedures

Once the final item parameters were estimated, they were used to create scoring tables. At this stage, the scoring tables produced by the final item calibration run in Winsteps are still on the theta metric. Each potential “number of total score points (multiple-choice items correct plus total points on constructed response items)” is associated with a theta estimate. This “person-level” theta would be their score on the theta scale. Output files were verified to match between HumRRO and Pearson at this stage.

Once theta scoring tables were obtained, they were linearly transformed to a reporting scale of 100-300 for all grade subjects. Performance levels (Novice, Apprentice, Proficient, and Distinguished) were also assigned to each score. Cut scores for the performance levels were determined following a standard setting workshop conducted in the summer of 2012 (see Pearson, 2012). The results of that workshop included cut scores on the theta metric that can be used to assign NAPD categories to students. Scale score cuts were used, as opposed to theta cuts, to assign performance levels to students’ scale scores. Using these cuts allowed the scale scores associated with each performance level to be fixed across test administrations. HumRRO verified the raw-score-to-scale-score tables and the associated performance levels.

In addition to overall scores, Kentucky also reports cluster scores (subscores based on subsets of items within each test). The generation of cluster scores uses the previously estimated item parameters and is accomplished by generating scoring tables in Winsteps on the theta metric, based on the specific items identified for each scoring cluster. These theta scores are then transformed in exactly the same manner as the full test scores.

Finally, for grade 5 Reading, grade 5 Mathematics and grade 7 Science there were items on the non-braille form that were unsuitable for braille examinees. Separate score tables for the braille form were produced and verified for these three tests. Separate theta estimates were calibrated for Braille, using the equating solution of the non-braille test, but omitting the unsuitable items. Using the common items, new score tables were constructed and compared to Pearson’s estimations.

Verification of 2013 Scoring Tables

After the final scoring tables were constructed, the scoring tables were applied to the 2013 student data. HumRRO checks the 2013 scored student data to verify that the scoring tables are being appropriately applied to the data and to check the distribution of students falling into each performance level. In addition to Reading, Mathematics, Science, and Social Studies, the performance level distribution was also verified for writing. HumRRO matched Pearson on the number and percent of students assigned to each performance level by subject and grade.

Documentation

As HumRRO and Pearson completed each step of the process described above, Winsteps control, item parameter, score, and output files were shared to check for inconsistencies. Winsteps output contained the number of cases in the calibration sample, item-level information (e.g., p-values, parameters), and the theta scoring tables. A sample of the output files are appended to this document. They include:

1. Winsteps Control Files (Appendix A). These files contain the item parameter estimation specifications and important information for reading the student score files. It also specifies the output file names. The appendix includes an example control file for the initial item parameter estimation, equated item parameter estimation, and estimation of the cluster scores.
2. Winstep Item Parameter Files (Appendix B). These files contain the item parameters for the operational items. Each multiple-choice item has one parameter, a logit difficulty (named Measure in the Wintsep files). Each constructed-response item has an overall difficulty parameter and a number of step parameters indicating how the points for the item are distributed along the theta scale. The file included in the appendix is an example of a final item parameter file. Initial item parameter files are in similar formats.
3. Winsteps Anchor File (Appendix D). The file includes the 2012 item parameter values for each anchor item. The file is read by Winsteps and used to fix the anchor item parameter values.
4. Winsteps Score File (Appendix E). The file contains the raw score to theta estimation and includes the distribution of student scores.
5. Comparison of Files Output (Appendix F). This is a SAS output file from HumRRO's comparison program that checks scoring table results against Pearson's results. The files match if all comparison values are 0.

Issues

This section of the report documents the issues that arose during the course of the replication and checking process that warrant some explanation/clarification. Both of the "issues" described below are simply matters that are worth noting for documentation purposes, but that do not require changes to procedures for subsequent years.

Earlier versions of the test specifications (before v1.5) contained two errors in the theta cuts presented in Table 15. The grade 4 reading cut for the Apprentice level and the grade 6 reading cut for the Distinguished level were not accurate. In these two instances, the theta cuts recommended by standard setting were used instead of the final KDE-approved theta cuts. This issue was identified after all analyses had been completed; thus the raw-score-to-scale-score tables and the impact data for these two grades had to be re-estimated and verified.

During the discussion on how to handle the first issue, KDE raised another concern. Earlier versions of the test specifications (before v1.5) stated that scale scores would be assigned performance levels using the theta cuts. KDE was concerned that this practice could confuse the general public since the scale scores associated with each performance level could vary slightly each year. KDE consulted Pearson, HumRRO, and EdMeasure to determine the best practice to use for the K-PREP. KDE decided to use the scale score cut point instead of the theta cut point so that the scores associated with each performance level would be consistent across test administrations. This decision required new raw-score-to-scale-score tables to be re-estimated and verified for each assessment.

Conclusion

Pearson and HumRRO independently calculated the scaled/equated raw-score-to-scale-score tables for the 2013 K-PREP assessments. No differences were found between Pearson's and HumRRO's parameter estimation, equating constants, or raw-score-to-scale-score tables. Given that HumRRO's and Pearson's scaling and equating results were identical, HumRRO is confident that Pearson did not commit processing errors.

References

- Bynum, B. H., & Thacker, A. A. (2011). *Third-party checking of calibration scaling and equating of the 2011 Kentucky core content test (FR-11-65)*. Alexandria, VA: Human Resources Research Organization.
- Bynum, B. H., & Thacker, A. A. (2013). *Third-party checking of 2012 scaling and equating of the for the Kentucky Performance Rating for Educational Progress (K-PREP) Tests (2013 No.11)*. Alexandria, VA: Human Resources Research Organization.
- Huynh, H. (2000, June). *Guidelines for Rasch Linking for PACT*. Memorandum to Paul Sandifer on June 18, 2000. Columbia, SC: Available from Author.
- Huynh, H., & Meyer, P. (2010). Use of robust z in detecting unstable items in item response theory models. *Practical Assessment, Research & Evaluation*, 15(2). Available online: <http://pareonline.net/getvn.asp?v=15&n=2>.
- Huynh, H., & Rawls, A. (2009). A comparison between robust z and 0.3-logit difference procedures in assessing stability of linking items for the Rasch model. In Everett V. Smith Jr. & Greg E. Stone (Eds.) *Applications of Rasch Measurement in Criterion-Referenced Testing: Practice Analysis to Score Reporting*. Maple Grove, MN: JAM Press.
- Thacker, A. A., Dickinson, E. R., & Sinclair, A. L. (2013). *Policy capture for setting end-of-course and Kentucky Performance Rating for Education Progress (K-PREP) cut scores (2013 No. 007)*. Alexandria, VA: Human Resources Research Organization.
- Pearson (2012). *Kentucky performance rating for educational progress performance standards workshop: Performance level descriptor creation and standard setting, v1.1*. Pearson, Inc.

Appendix A – Control File (Reading Grade 3)

```
;Winstep Control file - Initial parameter Estimation - a03RD_v0
; HumRRO
&INST
Item1 = 25
NI = 38
TABLES = 00100000000001000001000000001
CODES = 012
CSV = N
FITP = 3.0
FITI = 3.0
XWIDE =1
HLINES = Y
data=a03RDmopv0.dat
IFILE= a03RDv0.ITM
ISFILE = a03RDv0.ISF
SFILE = a03RDv0.CSF
;IAFILE = a03RDv0.IAF
;SAFILE = a03RDv0.SAF
SCFILE = a03RDv0.RSS
PFILE = a03RDv0.PER
mprox=10
mucon=100
rconv=.50
lconv=.01
models=r
groups=0
stkeep=n
realse=n
stbias=n
target=n
extrsc=0.25
udecim=4
uimean=0
uscale=1
;upmean=0
;uanchor=y
ptbis=y
ILFILES = *
2109676
2109677
2109697
2109698
2109701
2109705
2109706
2109707
2109709
2109711
2109718
```

2109721
201866
201869
201860
201870
201883
201877
201885
201880
201876
201878
202104
202103
201874
R3085
R3083
R3082
R3089
R3091
R3087
201772
201833
201839
201830
201832
201785
202205
*
&END
END NAMES

;Winstep Control file - Final Parameter Estimation - a03RD_v0
; HumRRO
&INST
Item1 = 25
NI = 38
TABLES = 0010000000001000001000000001
CODES = 012
CSV = N
FITP = 3.0
FITI = 3.0
XWIDE =1
HLINES = Y
data=a03RDmopv0.dat
IAFILE= a03RDv0anchors.IAF
SAFILE= a03RDv0anchors.SAF
IFILE= a03RDv0_eq.ITM
ISFILE = a03RDv0_eq.ISF
SFILE = a03RDv0_eq.CSF
SCFILE = a03RDv0_eq.RSS
PFILE = a03RDv0_eq.PER

```
mprox=10
mucon=100
rconv=.50
lconv=.01
models=r
groups=0
stkeep=n
realse=n
stbias=n
target=n
extrsc=0.25
udecim=4
uimean=0
uscale=1
;upmean=0
;uanchor=n
ptbis=y
ILFILES = *
2109676
2109677
2109697
2109698
2109701
2109705
2109706
2109707
2109709
2109711
2109718
2109721
201866
201869
201860
201870
201883
201877
201885
201880
201876
201878
202104
202103
201874
R3085
R3083
R3082
R3089
R3091
R3087
201772
201833
```

201839
201830
201832
201785
202205

*

&END
END NAMES

;Winstep Control file π Cluster Estimation RD1 - a03RD_v0

; HumRRO

&INST

Item1 = 25

NI = 38

TABLES = 001000000000010000010000000001

CODES = 012

CSV = N

FITP = 3.0

FITI = 3.0

XWIDE =1

HLINES = Y

data=a03RDmopv0.dat

IFILE= a03RDv0_RD1.ITM

ISFILE = a03RDv0_RD1.ISF

SFILE = a03RDv0_RD1.CSF

IAFILE = a03RDv0_eq.ITM

SAFILE = a03RDv0_eq.CSF

SCFILE = a03RDv0_RD1.RSS

PFILE = a03RDv0_RD1.PER

IDFILE = *

5

11

12

13

14

15

16

19

20

21

22

23

24

25

27

28

29

30

31

32

34
35
36
37
38
*
mprox=10
mucon=100
rconv=.50
lconv=.01
models=r
groups=0
stkeep=n
realse=n
stbias=n
target=n
extrsc=0.25
udecim=4
uimean=0
uscale=1
;upmean=0
;uanchor=y
ptbis=y
ILFILES = *
2109676
2109677
2109697
2109698
2109701
2109705
2109706
2109707
2109709
2109711
2109718
2109721
201866
201869
201860
201870
201883
201877
201885
201880
201876
201878
202104
202103
201874
R3085
R3083

R3082

R3089

R3091

R3087

201772

201833

201839

201830

201832

201785

202205

*

&END

END NAMES

Appendix B – Winsteps Item Parameter Files (Reading Grade 3)

```

; ITEM C:\Data\Kentucky\KPREP2013\WINSTEP\A03RDv0_eq.con Jul 11 16:24 2013
;ENTRY MEASURE ST COUNT SCORE ERROR IN.MSQ IN.ZST OUT.MS OUT.ZS DISPL PTBISE WEIGHT OBSMA EXPMA DISCRM LOWER UPPER PVALU PBE-E RMSR
1 -1.8273 1 50477.0 44725.0 .0149 .99 -.65 1.04 1.97 .0042 .27 1.00 89.0 89.0 1.00 .00 1.00 .89 .27 .30
2 -.5360 1 50477.0 36283.0 .0108 1.11 9.90 1.22 9.90 .0041 .23 1.00 72.9 75.1 .79 .09 .96 .72 .34 .43
3 -2.6164 1 50477.0 47453.0 .0198 .92 -4.88 .69 -9.90 .0049 .32 1.00 94.3 94.3 1.06 .00 1.00 .94 .22 .22
4 -.2064 1 50477.0 33314.0 .0103 .97 -6.57 .95 -6.74 .0040 .38 1.00 72.4 71.7 1.07 .00 1.00 .66 .35 .43
5 -1.0029 1 50477.0 39957.0 .0119 1.04 6.07 1.09 7.46 .0041 .27 1.00 80.1 80.5 .94 .00 .99 .79 .32 .38
6 1.0976 1 50477.0 20089.0 .0100 1.09 9.90 1.16 9.90 .0040 .24 1.00 65.4 69.1 .72 .06 .92 .40 .34 .47
7 -.3494 1 50477.0 34636.0 .0105 .88 -9.90 .82 -9.90 .0043 .48 1.00 77.1 73.1 1.25 .00 1.00 .69 .35 .40
8 -1.1204 1 50477.0 40768.0 .0122 .87 -9.90 .72 -9.90 .0053 .46 1.00 83.1 81.8 1.17 .00 1.00 .81 .31 .34
9 -1.3942 1 50477.0 42490.0 .0131 .85 -9.90 .67 -9.90 .0041 .47 1.00 85.9 84.8 1.17 .00 1.00 .84 .30 .31
10 -.4535 1 50477.0 35566.0 .0107 .92 -9.90 .86 -9.90 .0043 .43 1.00 76.1 74.2 1.16 .00 1.00 .70 .34 .40
11 -1.2065 1 50477.0 41338.0 .0125 .88 -9.90 .74 -9.90 .0047 .44 1.00 83.8 82.8 1.15 .00 1.00 .82 .31 .34
12 .0648 1 50477.0 30677.0 .0100 .98 -5.09 .96 -7.05 .0044 .38 1.00 69.8 69.4 1.06 .00 1.00 .61 .36 .44
13 -.5452 1 50477.0 36360.0 .0108 .99 -.98 .93 -7.61 .0045 .36 1.00 74.3 75.2 1.03 .07 1.00 .72 .34 .41
14 .9402 1 50477.0 21680.0 .0099 1.01 3.32 1.06 9.90 .0042 .32 1.00 68.9 68.2 .94 .04 1.00 .43 .35 .45
15 -1.0638 1 50477.0 40381.0 .0120 .94 -8.43 .89 -9.06 .0048 .38 1.00 82.2 81.2 1.07 .00 1.00 .80 .32 .36
16 .3376 1 50477.0 27912.0 .0099 1.00 1.13 1.00 -.10 .0047 .35 1.00 67.3 67.9 .99 .05 1.00 .55 .36 .45
17 .9662 2 50477.0 21465.0 .0099 1.12 9.90 1.19 9.90 -.0007 .22 1.00 63.2 68.3 .62 .09 .93 .43 .35 .48
18 .4711 2 50477.0 27682.0 .0098 1.01 2.39 1.01 1.40 -.1064 .34 1.00 67.1 67.5 .97 .06 1.00 .55 .36 .46
19 .0852 2 50477.0 30021.0 .0100 .95 -9.90 .92 -9.90 .0499 .42 1.00 70.7 69.3 1.15 .00 1.00 .59 .36 .43
20 .6012 2 50477.0 22460.0 .0098 1.04 9.90 1.06 9.90 .2659 .31 1.00 65.3 67.4 .87 .01 .95 .44 .36 .46
21 1.0512 1 50477.0 20555.0 .0100 .96 -9.90 .98 -3.51 .0043 .38 1.00 71.1 68.8 1.10 .00 1.00 .41 .35 .44
22 -.6088 1 50477.0 36899.0 .0109 .97 -4.92 .95 -5.40 .0041 .37 1.00 76.7 75.9 1.05 .00 1.00 .73 .34 .40
23 .1688 2 50477.0 30619.0 .0099 .99 -2.01 .98 -3.69 -.0940 .35 1.00 68.9 68.8 1.03 .06 1.00 .61 .36 .45
24 -1.1450 1 50477.0 40933.0 .0123 .83 -9.90 .68 -9.90 .0044 .50 1.00 84.2 82.1 1.21 .00 1.00 .81 .31 .33
25 1.0837 1 50477.0 39196.0 .0067 1.12 9.90 1.17 9.90 .0035 .38 1.00 48.3 51.9 .79 .09 1.97 .78 .46 .71
26 .6968 2 50477.0 25613.0 .0098 1.02 6.90 1.04 7.99 -.1320 .34 1.00 66.6 67.5 .91 .04 1.00 .51 .36 .46
27 .3602 2 50477.0 29683.0 .0099 1.06 9.90 1.07 9.90 -.1919 .28 1.00 65.2 67.8 .81 .18 1.00 .59 .36 .47
28 .3049 2 50477.0 30787.0 .0099 .96 -9.90 .95 -9.62 -.2468 .38 1.00 70.0 68.1 1.12 .07 1.00 .61 .36 .44
29 1.7107 1 50477.0 14336.0 .0108 1.25 9.90 1.53 9.90 .0038 .06 1.00 69.0 74.8 .50 .10 .68 .28 .32 .46
30 -1.1982 1 50477.0 41284.0 .0124 .91 -9.90 .80 -9.90 .0038 .41 1.00 83.3 82.7 1.11 .00 1.00 .82 .31 .34
31 -1.1049 1 50477.0 40665.0 .0121 .93 -9.70 .85 -9.90 .0044 .39 1.00 82.4 81.6 1.09 .00 1.00 .81 .31 .35
32 -1.4448 1 50477.0 42782.0 .0133 .96 -4.73 .98 -1.06 .0051 .33 1.00 86.0 85.3 1.03 .00 1.00 .85 .29 .33
33 .3647 2 50477.0 26686.0 .0099 1.11 9.90 1.15 9.90 .0957 .26 1.00 63.2 67.8 .66 .08 .92 .53 .36 .48
34 .2940 2 50477.0 27334.0 .0099 1.09 9.90 1.12 9.90 .1038 .27 1.00 63.7 68.1 .72 .07 .94 .54 .36 .47
35 -.4302 2 50477.0 36295.0 .0106 .91 -9.90 .87 -9.90 -.1046 .40 1.00 76.4 74.0 1.17 .00 1.00 .72 .34 .40
36 -1.1163 2 50477.0 39829.0 .0122 1.00 .07 .97 -2.45 .1380 .39 1.00 81.5 81.8 1.01 .00 1.00 .79 .31 .37
37 -.3031 1 50477.0 34212.0 .0104 1.11 9.90 1.16 9.90 .0044 .23 1.00 68.9 72.6 .76 .12 .95 .68 .35 .45
38 .8021 1 50477.0 45403.0 .0064 1.05 8.57 1.05 7.13 .0040 .45 1.00 48.5 49.3 .93 .02 1.99 .90 .48 .71

```

Step Parameters

; STRUCTURE MEASURE C:\Data\Kentucky\KPREP2013\WINSTEP\a03RDv0_eq.con Jul 11
16:24 2013

; ITEM CATEGORY Rasch-Andrich threshold MEASURE

1	0	.0000
1	1	.0000
2	0	.0000
2	1	.0000
3	0	.0000
3	1	.0000
4	0	.0000
4	1	.0000
5	0	.0000
5	1	.0000
6	0	.0000
6	1	.0000
7	0	.0000
7	1	.0000
8	0	.0000
8	1	.0000
9	0	.0000
9	1	.0000
10	0	.0000
10	1	.0000
11	0	.0000
11	1	.0000
12	0	.0000
12	1	.0000
13	0	.0000
13	1	.0000
14	0	.0000
14	1	.0000
15	0	.0000
15	1	.0000
16	0	.0000
16	1	.0000
17	0	.0000
17	1	.0000
18	0	.0000
18	1	.0000
19	0	.0000
19	1	.0000
20	0	.0000
20	1	.0000
21	0	.0000
21	1	.0000
22	0	.0000
22	1	.0000
23	0	.0000
23	1	.0000

24	0	.0000
24	1	.0000
25	0	.0000
25	1	-.2203
25	2	.2203
26	0	.0000
26	1	.0000
27	0	.0000
27	1	.0000
28	0	.0000
28	1	.0000
29	0	.0000
29	1	.0000
30	0	.0000
30	1	.0000
31	0	.0000
31	1	.0000
32	0	.0000
32	1	.0000
33	0	.0000
33	1	.0000
34	0	.0000
34	1	.0000
35	0	.0000
35	1	.0000
36	0	.0000
36	1	.0000
37	0	.0000
37	1	.0000
38	0	.0000
38	1	.0284
38	2	-.0284

Appendix C – Robust Z Item Stability Analysis

- Step 1.** Calculate the mean and standard deviation of the 2012 item parameter estimates of the linking items: b-parameter for multiple-choice items and step parameter estimates for open-ended items.
- Step 2.** Calculate the mean and standard deviation of the 2013 item parameter estimates of the linking items: b-parameter for multiple-choice items and step parameter estimates for open-ended items.
- Step 3.** Calculate the ratio of standard deviations.
- Step 4.** Calculate the correlation between the 2012 and 2013 item parameter estimates of the linking items.
- Step 5.** Calculate the difference between the 2012 and 2013 item parameter estimates for each linking item (e.g., $b_{2012}-b_{2013}$ for multiple-choice item; $d_{1,2012}-d_{1,2013}$ for open-ended item).
- Step 6.** Calculate the median and interquartile range of the differences calculated in Step 5.
- Step 7.** Calculate the robust z statistic for each linking item using the following equation:

$$Z = \frac{D - M_d}{0.74 * IQR},$$

where D is the difference in item parameter estimates, M_d is the median of the differences, and IQR is the interquartile range of the differences.

Once all robust z statistics have been calculated for a linking set, the following guidelines dictate removing linking items from equating:

- a. The ratio of standard deviations must be in the 0.9-1.1 range; the correlation of item parameter estimates must be greater than 0.95.
- b. If either condition in (a) is not met, then remove the linking item with the largest absolute robust z value (assuming it is greater than 1.645). *Note: If one step difficulty of the open-ended item is removed then all other step difficulties for that open-ended item are removed as well.*
- c. Repeat steps 1 through 4 – do not recalculate robust z statistics – and continue removing linking items until:
 - the standard deviation ratio and correlation of item difficulties are within the prescribed range; or
 - there are no linking items with an absolute robust z value greater than 1.645; or
 - 20% of the linking item set has been removed.³

³ This will be discussed if the number of flagged items for removal exceeds the criterion.

Appendix D – Winsteps Anchor File (Grade 3 Reading)

Multiple Choice Item Anchor File

ENTRY	MEASURE	ST	COUNT	SCORE	ERROR	IN.MSQ	IN.ZST	OUT.MS	OUT.ZS	DISPL	PTBISE	WEIGHT	OBSMA	EXPMA	DISCRM	LOWER	UPPER	PVALU	PBE-E	RMSR
30	1.7959	1	50477.0	15498.0	.0109	1.16	9.90	1.39	9.90	.0003	.24	1.00	72.5	75.6	.68	.07	.94	.31	.40	.44
14	-.1890	1	50477.0	35690.0	.0108	1.01	2.58	1.03	2.53	-.0002	.37	1.00	74.2	74.5	.97	.00	.99	.71	.39	.41
40	-.1821	1	50477.0	35370.0	.0108	1.01	2.26	.98	-2.08	-.0003	.38	1.00	73.6	74.2	.99	.03	1.00	.70	.39	.42
19	.2751	1	50477.0	30925.0	.0102	.91	-9.90	.87	-9.90	-.0001	.49	1.00	74.4	70.8	1.23	.00	1.00	.61	.41	.42
33	1.7092	1	50477.0	37528.0	.0093	.95	-8.71	.95	-8.60	.0011	.49	1.00	69.5	68.7	1.07	.00	2.00	.74	.44	.47
18	2.0391	1	50477.0	15374.0	.0109	.96	-7.87	1.05	4.98	.0004	.40	1.00	78.0	75.7	1.04	.00	1.00	.30	.40	.40
12	.2489	1	50477.0	29786.0	.0102	1.04	9.21	1.06	7.63	-.0001	.38	1.00	68.7	70.3	.90	.03	.98	.59	.41	.45
25	1.3312	1	50477.0	21518.0	.0102	1.09	9.90	1.14	9.90	.0001	.33	1.00	67.7	71.0	.77	.06	.97	.43	.42	.46
11	-.9430	1	50477.0	40143.0	.0120	.92	-9.90	.87	-9.12	-.0004	.43	1.00	81.7	80.6	1.11	.00	1.00	.80	.35	.36
38	1.8666	1	50477.0	15160.0	.0110	1.11	9.90	1.31	9.90	.0003	.28	1.00	74.5	76.0	.77	.05	.96	.30	.40	.43
10	.7838	1	50477.0	27647.0	.0101	.93	-9.90	.89	-9.90	.0000	.48	1.00	72.7	69.8	1.19	.00	1.00	.55	.42	.43
17	-.5155	1	50477.0	39756.0	.0119	.96	-6.30	.93	-5.22	-.0003	.39	1.00	80.9	79.9	1.05	.00	1.00	.79	.35	.37
16	-.1486	1	50477.0	34025.0	.0106	.88	-9.90	.82	-9.90	-.0004	.51	1.00	76.8	72.9	1.26	.00	1.00	.67	.40	.40
28	.2166	1	50477.0	29281.0	.0101	.97	-7.58	1.00	.43	.0000	.44	1.00	72.2	70.1	1.06	.00	1.00	.58	.42	.43
27	-.0754	1	50477.0	33929.0	.0105	1.00	-.81	.95	-5.40	-.0001	.40	1.00	72.2	72.8	1.02	.03	1.00	.67	.40	.42
35	.5853	1	50477.0	28376.0	.0101	1.06	9.90	1.07	9.90	.0000	.36	1.00	67.1	69.9	.83	.06	.98	.56	.42	.46
45	-.7340	1	50477.0	36822.0	.0110	.95	-9.90	.91	-8.49	-.0002	.43	1.00	77.4	75.8	1.09	.00	1.00	.73	.38	.39
37	.3416	1	50477.0	30502.0	.0102	.94	-9.90	.91	-9.90	.0001	.46	1.00	72.8	70.6	1.15	.00	1.00	.60	.41	.42
22	1.0913	1	50477.0	23792.0	.0101	.95	-9.90	.96	-5.75	-.0001	.46	1.00	73.1	70.1	1.11	.00	1.00	.47	.42	.43
46	-.1561	1	50477.0	34372.0	.0106	.94	-9.90	.91	-9.19	.0001	.45	1.00	75.1	73.2	1.12	.00	1.00	.68	.39	.41
31	1.0848	1	50477.0	22727.0	.0101	1.11	9.90	1.17	9.90	.0001	.32	1.00	66.5	70.4	.71	.07	.94	.45	.42	.46
47	.4518	1	50477.0	61454.0	.0070	1.02	3.05	1.06	7.27	.0000	.53	1.00	54.6	54.6	.96	.00	1.97	1.22	.54	.64
41	.2926	1	50477.0	29841.0	.0102	.93	-9.90	.89	-9.90	-.0001	.48	1.00	73.1	70.3	1.18	.00	1.00	.59	.41	.42
34	.2186	1	50477.0	30295.0	.0102	.88	-9.90	.83	-9.90	-.0002	.52	1.00	75.2	70.5	1.30	.00	1.00	.60	.41	.41

Step Parameter Anchor File

; ITEM CATEGORY Rasch-Andrich threshold MEASURE

10 0 0.0000
 10 1 0.0000
 11 0 0.0000
 11 1 0.0000
 12 0 0.0000
 12 1 0.0000
 14 0 0.0000
 14 1 0.0000
 16 0 0.0000
 16 1 0.0000
 17 0 0.0000
 17 1 0.0000

18	0	0.0000
18	1	0.0000
19	0	0.0000
19	1	0.0000
22	0	0.0000
22	1	0.0000
25	0	0.0000
25	1	0.0000
27	0	0.0000
27	1	0.0000
28	0	0.0000
28	1	0.0000
30	0	0.0000
30	1	0.0000
31	0	0.0000
31	1	0.0000
33	0	0.0000
33	1	-2.1157
33	2	2.1157
34	0	0.0000
34	1	0.0000
35	0	0.0000
35	1	0.0000
37	0	0.0000
37	1	0.0000
38	0	0.0000
38	1	0.0000
40	0	0.0000
40	1	0.0000
41	0	0.0000
41	1	0.0000
45	0	0.0000
45	1	0.0000
46	0	0.0000
46	1	0.0000
47	0	0.0000
47	1	-0.1802
47	2	0.1801

Appendix E – Winsteps Score File (Grade 3 Reading)

PERSON SCORE FILE FOR C:\Data\Kentucky\KPREP2013\WINSTEP\A03MAV0_eq.con Jul 10 10:02 2013 USCALE=1.00

SCORE	MEASURE	S.E.	INFO	NORMED	S.E.	FREQUENCY	%	CUM.FREQ.	%	PERCENTILE
0	-5.5619	2.0095	.25	-21	163	81	.2	81	.2	1
1	-4.1469	1.0190	.96	94	83	34	.1	115	.2	1
2	-3.4153	.7339	1.86	153	60	4	.0	119	.2	1
3	-2.9711	.6101	2.69	189	50	6	.0	125	.2	1
4	-2.6443	.5378	3.46	216	44	5	.0	130	.3	1
5	-2.3818	.4895	4.17	237	40	12	.0	142	.3	1
6	-2.1597	.4545	4.84	255	37	26	.1	168	.3	1
7	-1.9654	.4280	5.46	271	35	39	.1	207	.4	1
8	-1.7913	.4072	6.03	285	33	73	.1	280	.6	1
9	-1.6325	.3903	6.56	298	32	167	.3	447	.9	1
10	-1.4857	.3765	7.05	310	31	206	.4	653	1.3	1
11	-1.3483	.3650	7.51	321	30	287	.6	940	1.9	2
12	-1.2187	.3552	7.92	332	29	403	.8	1343	2.7	2
13	-1.0955	.3470	8.31	342	28	471	.9	1814	3.6	3
14	-.9776	.3399	8.66	351	28	653	1.3	2467	4.9	4
15	-.8642	.3339	8.97	361	27	746	1.5	3213	6.4	6
16	-.7545	.3287	9.26	370	27	859	1.7	4072	8.1	7
17	-.6480	.3242	9.51	378	26	977	1.9	5049	10.0	9
18	-.5442	.3204	9.74	387	26	1042	2.1	6091	12.1	11
19	-.4426	.3171	9.94	395	26	1096	2.2	7187	14.2	13
20	-.3429	.3144	10.11	403	26	1261	2.5	8448	16.7	15
21	-.2447	.3122	10.26	411	25	1318	2.6	9766	19.3	18
22	-.1479	.3104	10.38	419	25	1304	2.6	11070	21.9	21
23	-.0519	.3091	10.47	427	25	1378	2.7	12448	24.7	23
24	.0433	.3081	10.53	434	25	1504	3.0	13952	27.6	26
25	.1380	.3076	10.57	442	25	1522	3.0	15474	30.7	29
26	.2326	.3075	10.58	450	25	1559	3.1	17033	33.7	32
27	.3272	.3078	10.56	458	25	1561	3.1	18594	36.8	35
28	.4221	.3085	10.50	465	25	1610	3.2	20204	40.0	38
29	.5177	.3098	10.42	473	25	1659	3.3	21863	43.3	42
30	.6142	.3116	10.30	481	25	1654	3.3	23517	46.6	45
31	.7120	.3139	10.15	489	26	1644	3.3	25161	49.8	48
32	.8114	.3170	9.95	497	26	1693	3.4	26854	53.2	52
33	.9130	.3207	9.72	505	26	1637	3.2	28491	56.4	55

34	1.0174	.3253	9.45	514	26	1646	3.3	30137	59.7	58
35	1.1250	.3309	9.13	522	27	1679	3.3	31816	63.0	61
36	1.2367	.3376	8.77	532	27	1719	3.4	33535	66.4	65
37	1.3533	.3456	8.37	541	28	1654	3.3	35189	69.7	68
38	1.4760	.3551	7.93	551	29	1777	3.5	36966	73.2	71
39	1.6060	.3664	7.45	562	30	1735	3.4	38701	76.7	75
40	1.7452	.3800	6.93	573	31	1737	3.4	40438	80.1	78
41	1.8956	.3962	6.37	585	32	1592	3.2	42030	83.3	82
42	2.0602	.4159	5.78	599	34	1549	3.1	43579	86.3	85
43	2.2430	.4401	5.16	613	36	1421	2.8	45000	89.1	88
44	2.4499	.4706	4.51	630	38	1294	2.6	46294	91.7	90
45	2.6897	.5104	3.84	650	42	1241	2.5	47535	94.2	93
46	2.9771	.5644	3.14	673	46	1034	2.0	48569	96.2	95
47	3.3390	.6433	2.42	703	52	848	1.7	49417	97.9	97
48	3.8335	.7738	1.67	743	63	593	1.2	50010	99.1	98
49	4.6380	1.0610	.89	808	86	344	.7	50354	99.8	99
50	6.1279	2.0374	.24	929	166	123	.2	50477	100.0	99

Appendix F – Comparison of Files Output (Reading Grade 3)

a03RDv0_eq ITEM Parameter comparison

Obs	rs	theta	theta_hum	se	SE_Hum	ss	SS_hum	PL	PL_hum	theta_diff	SS_diff	PL_diff
1	0	-5.6956	-5.6956	2.0156	2.0156	104	104	N	N	0	0	0
2	1	-4.2631	-4.2631	1.03	1.03	127	127	N	N	0	0	0
3	2	-3.5101	-3.5101	0.7476	0.7476	140	140	N	N	0	0	0
4	3	-3.0464	-3.0464	0.6253	0.6253	148	148	N	N	0	0	0
5	4	-2.7014	-2.7014	0.554	0.554	153	153	N	N	0	0	0
6	5	-2.4217	-2.4217	0.5064	0.5064	158	158	N	N	0	0	0
7	6	-2.183	-2.183	0.4722	0.4722	162	162	N	N	0	0	0
8	7	-1.9725	-1.9725	0.4464	0.4464	166	166	N	N	0	0	0
9	8	-1.7825	-1.7825	0.4263	0.4263	169	169	N	N	0	0	0
10	9	-1.6078	-1.6078	0.4102	0.4102	172	172	N	N	0	0	0
11	10	-1.445	-1.445	0.3972	0.3972	174	174	N	N	0	0	0
12	11	-1.2915	-1.2915	0.3866	0.3866	177	177	N	N	0	0	0
13	12	-1.1456	-1.1456	0.3778	0.3778	179	179	N	N	0	0	0
14	13	-1.0057	-1.0057	0.3705	0.3705	182	182	N	N	0	0	0
15	14	-0.8707	-0.8707	0.3645	0.3645	184	184	N	N	0	0	0
16	15	-0.7398	-0.7398	0.3595	0.3595	186	186	N	N	0	0	0
17	16	-0.612	-0.612	0.3555	0.3555	188	188	N	N	0	0	0
18	17	-0.4868	-0.4868	0.3523	0.3523	190	190	N	N	0	0	0
19	18	-0.3637	-0.3637	0.3498	0.3498	192	192	N	N	0	0	0
20	19	-0.242	-0.242	0.348	0.348	194	194	N	N	0	0	0
21	20	-0.1213	-0.1213	0.3469	0.3469	196	196	N	N	0	0	0
22	21	-0.0012	-0.0012	0.3464	0.3464	198	198	A	A	0	0	0
23	22	0.1188	0.1188	0.3466	0.3466	200	200	A	A	0	0	0
24	23	0.2392	0.2392	0.3475	0.3475	202	202	A	A	0	0	0
25	24	0.3605	0.3605	0.3491	0.3491	204	204	A	A	0	0	0
26	25	0.4832	0.4832	0.3516	0.3516	207	207	A	A	0	0	0
27	26	0.608	0.608	0.3551	0.3551	209	209	A	A	0	0	0
28	27	0.7356	0.7356	0.3596	0.3596	211	211	P	P	0	0	0
29	28	0.867	0.867	0.3656	0.3656	213	213	P	P	0	0	0
30	29	1.0033	1.0033	0.3731	0.3731	215	215	P	P	0	0	0
31	30	1.1459	1.1459	0.3826	0.3826	218	218	P	P	0	0	0
32	31	1.2967	1.2967	0.3945	0.3945	220	220	P	P	0	0	0

33	32	1.4582	1.4582	0.4097	0.4097	223	223	P	P	0	0	0
34	33	1.6338	1.6338	0.4292	0.4292	226	226	D	D	0	0	0
35	34	1.8286	1.8286	0.4546	0.4546	229	229	D	D	0	0	0
36	35	2.0503	2.0503	0.4887	0.4887	233	233	D	D	0	0	0
37	36	2.3117	2.3117	0.5365	0.5365	237	237	D	D	0	0	0
38	37	2.6368	2.6368	0.6086	0.6086	242	242	D	D	0	0	0
39	38	3.079	3.0791	0.7325	0.7325	250	250	D	D	-0.0001	0	0
40	39	3.8086	3.8086	1.018	1.018	262	262	D	D	0	0	0
41	40	5.2222	5.2222	2.0091	2.0091	286	286	D	D	0	0	0