



Third-Party Checking of 2015 Scaling and Equating for the Kentucky Performance Rating for Educational Progress (K-PREP) Tests

Final Report

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

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

Prepared under: Human Resources Research Organization
700 North Hurstbourne Parkway, Suite 100
Louisville, KY 40222

Date: March 22, 2016

Third-Party Checking of 2015 Scaling and Equating for the Kentucky Performance Rating for Educational Progress (K-PREP) Tests

Table of Contents

Executive Summary	ii
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 2015 Scoring Tables	3
Documentation.....	4
Conclusion	4
References.....	5
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 2015 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, July 17, 2015). 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 2015 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 Item-Response Theory (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.

Equating was added to the process in 2013 to permit comparison of the results across test years. The 2014 tests were equated to the 2013 tests using linking items. In this manner, comparable scores were produced for the 2014 K-PREP. This process was continued in 2015. In addition, scale scores were computed for the On-Demand Writing tests where simple number correct scores had been used in the past².

This report describes how student test responses for the 2015 K-PREP assessments were used to create scale scores and place students in Novice, Apprentice, Proficient or Advanced (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 Pearson³ and regularly updated based on decisions before and during calibration. This documentation is vital for ensuring consistency of processing across years and serves as 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

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

² For additional information on how writing was calibrated and scaled, see K-PREP ODW Calibration and Scaling Specs v0.4.docx.

³ K-PREP Sp15 CES Specs_KDE v1.2.docx.

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 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. Kentucky uses common forms for all operational items (those that contribute to student scores), but the forms differ on field-test items. Field-test items are used to build future forms.

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 2015, we equated to the prior scale for Reading and grade 8 Math⁴. 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. . K-PREP also has embedded SAT-10 items, some of which were included as anchor items. Both multiple-choice and short answer 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 prior year item difficulty estimates with the 2015 item difficulty estimates for the linking items (*b*-parameter estimates for multiple-choice items and step parameters for open-ended items) to identify items aberrant changes. See Appendix C for an outline of the steps in the procedure and the criteria for exclusion. Any items

⁴ We did not conduct equating analyses for Math grades 3-7, Science and Social Studies; prior scoring tables were applied.

flagged for removal were discussed with Pearson at this stage. Lastly, a final calibration was run using the prior year item parameter estimates as anchor values for the linking items, thereby placing the 2015 tests onto the same measurement scale as prior year tests. . The multiple-choice items were anchored to the prior year item Rasch difficulty parameter estimates and the open-ended response items were anchored to both the prior year 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.

For reading grades 3, 4, and 6 and math grades 3-8, the overall model criteria for removing linking items (see Appendix C for specific criteria) was within acceptable ranges, suggesting no need to remove any items from equating. For reading grades 5, 7 and 8, two, two and one item, respectively, flagged as showing aberrant changes between the prior year and current year parameter estimation. As a result, those items were removed from the anchor set.

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, 2013). 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 grades 3-7 Reading and grade5 Math 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 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 2014 Scoring Tables

After the final scoring tables were constructed, the scoring tables were applied to the 2015 student data. HumRRO checks the 2015 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. HumRRO verified Reading, Math, Science, Social Studies, and Writing

performance level distributions. 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 Winstep 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 2012-13 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.

Conclusion

Pearson and HumRRO independently calculated the scaled/equated raw-score-to-scale-score tables for the 2015 K-PREP assessments. No differences were found between Pearson's and HumRRO's parameter estimation, Stocking-Lord transformation 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 d03RD_v0
; HumRRO
&INST
Item1 = 25
NI = 38
TABLES = 001000000000010000010000000001
CODES = 012
CSV = N
FITP = 3.0
FITI = 3.0
XWIDE =1
HLINES = Y
data=d03RDmopv0.dat
IFILE= d03RDv0.ITM
ISFILE = d03RDv0.ISF
SFILE = d03RDv0.CSF
;IAFILE = d03RDv0.IAF
;SAFILE = d03RDv0.SAF
SCFILE = d03RDv0.RSS
PFILE = d03RDv0.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
R3022
R3023
R3027
R3030
R3024
201885
201880
201876
201878
202104
202103
201874
R3083
R3082
R3089
R3091
R3087
R3251
R3262
R3253
R3254
R3258
R3257
R3250
R3255
R3263
*
&END
END NAMES

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

```

; ITEM STRUCTURE FILE (not for anchoring: use SFILE=) FOR D:\Data\Kentucky\KPREP2015\WINSTEP\d03RDv0_eq.con Jul 10 11:35 2015
;ENTRY STAT MAX CAT BOT+.25 CAT STRU MEASURE ERROR CAT-0.5 AT CAT 50%PRB CAT STRU MEASURE ERROR CAT-0.5 TOP-.25 50%PRB
1 1 1 0 -2.8478 1 1 -1.7492 .0150 -1.7492 -.6506 -1.7492 0 0 .0000 .0000 .0000 .0000 .0000
2 1 1 0 -1.5382 1 1 -.4396 .0109 -.4396 .6591 -.4396 0 0 .0000 .0000 .0000 .0000 .0000
3 1 1 0 -3.6006 1 1 -2.5020 .0196 -2.5020 -1.4034 -2.5020 0 0 .0000 .0000 .0000 .0000 .0000
4 2 1 0 -1.3050 1 1 -.2064 .0105 -.2064 .8922 -.2064 0 0 .0000 .0000 .0000 .0000 .0000
5 1 1 0 -2.1319 1 1 -1.0332 .0123 -1.0332 .0654 -1.0332 0 0 .0000 .0000 .0000 .0000 .0000
6 2 1 0 -.0010 1 1 1.0976 .0099 1.0976 2.1962 1.0976 0 0 .0000 .0000 .0000 .0000 .0000
7 2 1 0 -1.4480 1 1 -.3494 .0107 -.3494 .7492 -.3494 0 0 .0000 .0000 .0000 .0000 .0000
8 1 1 0 -2.1430 1 1 -1.0444 .0123 -1.0444 .0543 -1.0444 0 0 .0000 .0000 .0000 .0000 .0000
9 2 1 0 -2.4928 1 1 -1.3942 .0135 -1.3942 -.2956 -1.3942 0 0 .0000 .0000 .0000 .0000 .0000
10 2 1 0 -1.5521 1 1 -.4535 .0109 -.4535 .6451 -.4535 0 0 .0000 .0000 .0000 .0000 .0000
11 1 1 0 -2.2525 1 1 -1.1539 .0126 -1.1539 -.0553 -1.1539 0 0 .0000 .0000 .0000 .0000 .0000
12 1 1 0 -.9032 1 1 .1954 .0100 .1954 1.2940 .1954 0 0 .0000 .0000 .0000 .0000 .0000
13 1 1 0 -1.4305 1 1 -.3319 .0107 -.3319 .7668 -.3319 0 0 .0000 .0000 .0000 .0000 .0000
14 1 1 0 .2636 1 1 1.3622 .0101 1.3622 2.4608 1.3622 0 0 .0000 .0000 .0000 .0000 .0000
15 1 1 0 -.8940 1 1 .2046 .0100 .2046 1.3032 .2046 0 0 .0000 .0000 .0000 .0000 .0000
16 1 1 0 -1.3475 1 1 -.2489 .0105 -.2489 .8497 -.2489 0 0 .0000 .0000 .0000 .0000 .0000
17 1 1 0 -1.0648 1 1 .0338 .0102 .0338 1.1325 .0338 0 0 .0000 .0000 .0000 .0000 .0000
18 2 1 0 -1.0134 1 1 .0852 .0101 .0852 1.1838 .0852 0 0 .0000 .0000 .0000 .0000 .0000
19 2 1 0 -.4974 1 1 .6012 .0098 .6012 1.6998 .6012 0 0 .0000 .0000 .0000 .0000 .0000
20 2 1 0 -.0474 1 1 1.0512 .0099 1.0512 2.1498 1.0512 0 0 .0000 .0000 .0000 .0000 .0000
21 1 1 0 -1.5815 1 1 -.4828 .0109 -.4828 .6158 -.4828 0 0 .0000 .0000 .0000 .0000 .0000
22 2 1 0 -.9298 1 1 .1688 .0101 .1688 1.2674 .1688 0 0 .0000 .0000 .0000 .0000 .0000
23 1 1 0 -2.1963 1 1 -1.0977 .0125 -1.0977 .0009 -1.0977 0 0 .0000 .0000 .0000 .0000 .0000
24 2 2 0 -.5474 1 1 .8634 .0105 .1819 1.0837 .4954 2 2 1.3040 .0116 1.9854 2.7148 1.6721
25 2 1 0 -.7384 1 1 .3602 .0099 .3602 1.4588 .3602 0 0 .0000 .0000 .0000 .0000 .0000
26 2 1 0 -.7937 1 1 .3049 .0100 .3049 1.4035 .3049 0 0 .0000 .0000 .0000 .0000 .0000
27 1 1 0 .8065 1 1 1.9051 .0109 1.9051 3.0037 1.9051 0 0 .0000 .0000 .0000 .0000 .0000
28 2 1 0 -2.2968 1 1 -1.1982 .0128 -1.1982 -.0996 -1.1982 0 0 .0000 .0000 .0000 .0000 .0000
29 1 1 0 -2.0112 1 1 -.9126 .0119 -.9126 .1860 -.9126 0 0 .0000 .0000 .0000 .0000 .0000
30 1 1 0 -.5247 1 1 .5739 .0098 .5739 1.6726 .5739 0 0 .0000 .0000 .0000 .0000 .0000
31 2 1 0 -.8624 1 1 .2362 .0100 .2362 1.3348 .2362 0 0 .0000 .0000 .0000 .0000 .0000
32 2 1 0 -1.7580 1 1 -.6594 .0113 -.6594 .4392 -.6594 0 0 .0000 .0000 .0000 .0000 .0000
33 1 1 0 -2.1281 1 1 -1.0295 .0122 -1.0295 .0692 -1.0295 0 0 .0000 .0000 .0000 .0000 .0000
34 1 1 0 .3275 1 1 1.4261 .0102 1.4261 2.5248 1.4261 0 0 .0000 .0000 .0000 .0000 .0000
35 2 1 0 -.5875 1 1 .5111 .0099 .5111 1.6097 .5111 0 0 .0000 .0000 .0000 .0000 .0000
36 2 1 0 -.1250 1 1 .9736 .0099 .9736 2.0722 .9736 0 0 .0000 .0000 .0000 .0000 .0000
37 2 1 0 -1.2452 1 1 -.1466 .0104 -.1466 .9520 -.1466 0 0 .0000 .0000 .0000 .0000 .0000
38 2 2 0 -1.7307 1 1 -.5389 .0122 -.8125 .4516 -.6545 2 2 1.4421 .0108 1.7157 2.6340 1.5576

```

Step Parameters

```
; STRUCTURE MEASURE ANCHOR FILE FOR  
D:\Data\Kentucky\KPREP2015\WINSTEP\d03RDv0.con Jul 8 11:54 2015  
; 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	-.3865
24	2	.3865
25	0	.0000
25	1	.0000
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	-.8892
38	2	.8892

Appendix C – Robust Z Item Stability Analysis

- Step 1.** Calculate the mean and standard deviation of the previous 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 2015 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 2015 and previous item parameter estimates of the linking items.
- Step 5.** Calculate the difference between the 2015 and previous item parameter estimates for each linking item (e.g., $b_{old}-b_{2014}$ for multiple-choice item; $d_{1,old}-d_{1,2014}$ 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 (SAS default: Definition 5).

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.⁵

⁵ 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
24	1.0837	1	50970.0	38626.0	.0069	1.10	9.90	1.16	9.90	.0003	.39	1.00	50.2	53.8	.46	.68	0	R	.201874	
20	1.0512	1	50970.0	21521.0	.0099	.97	-6.72	1.00	.85	.0001	.38	1.00	70.8	68.9	.36	.44	0	R	.201876	
19	.6012	1	50970.0	24839.0	.0098	1.08	9.90	1.12	9.90	-.0001	.29	1.00	64.9	68.1	.37	.47	0	R	.201880	
18	.0852	1	50970.0	31805.0	.0101	.94	-9.90	.90	-9.90	-.0004	.43	1.00	72.5	70.7	.37	.43	0	R	.201885	
22	.1688	1	50970.0	31134.0	.0101	1.02	6.00	1.01	1.64	-.0006	.35	1.00	68.8	70.2	.38	.45	0	R	.202104	
4	-.2064	1	50970.0	33731.0	.0103	.97	-6.16	.96	-5.20	-.0006	.40	1.00	73.3	72.4	.37	.42	0	R	.2109698	
6	1.0976	1	50970.0	21081.0	.0100	1.11	9.90	1.18	9.90	.0000	.26	1.00	65.3	69.2	.36	.47	0	R	.2109705	
7	-.3494	1	50970.0	35035.0	.0105	.88	-9.90	.82	-9.90	-.0003	.49	1.00	77.3	73.7	.37	.40	0	R	.2109706	
9	-1.3942	1	50970.0	42418.0	.0128	.84	-9.90	.66	-9.90	-.0011	.50	1.00	85.2	84.0	.32	.32	0	R	.2109709	
10	-.4535	1	50970.0	35699.0	.0107	.93	-9.90	.87	-9.90	-.0004	.44	1.00	76.0	74.4	.36	.40	0	R	.2109711	
26	.3049	1	50970.0	32690.0	.0102	.98	-3.94	.95	-7.73	-.0001	.39	1.00	71.3	71.4	.37	.43	0	R	.R3082	
25	.3602	1	50970.0	29113.0	.0099	1.19	9.90	1.24	9.90	-.0001	.19	1.00	60.4	69.1	.38	.49	0	R	.R3083	
28	-1.1982	1	50970.0	41940.0	.0126	.89	-9.90	.76	-9.90	-.0011	.45	1.00	84.3	83.2	.33	.33	0	R	.R3091	
36	.9736	1	50970.0	22377.0	.0099	1.11	9.90	1.18	9.90	.0001	.25	1.00	64.7	68.6	.37	.47	0	R	.R3250	
32	-.6594	1	50970.0	39806.0	.0117	.99	-1.80	1.01	.57	-.0006	.35	1.00	80.4	79.8	.34	.38	0	R	.R3253	
37	-.1466	1	50970.0	36518.0	.0108	1.00	-.83	1.00	-.35	-.0001	.37	1.00	75.4	75.4	.36	.41	0	R	.R3255	
35	.5111	1	50970.0	31766.0	.0101	.91	-9.90	.88	-9.90	-.0001	.46	1.00	74.2	70.7	.37	.42	0	R	.R3257	
31	.2362	1	50970.0	30447.0	.0100	.97	-7.55	.95	-8.88	-.0001	.40	1.00	70.7	69.8	.38	.44	0	R	.R3262	
38	.4516	1	50970.0	59287.0	.0073	.91	-9.90	.90	-9.90	-.0005	.54	1.00	58.1	55.5	.47	.58	0	R	.R3263	

Step Parameter Anchor File

```

; ITEM CATEGORY Rasch-Andrich threshold MEASURE
4 0 0.0000
4 1 0.0000
6 0 0.0000
6 1 0.0000
7 0 0.0000
7 1 0.0000
9 0 0.0000
9 1 0.0000
10 0 0.0000
10 1 0.0000
18 0 0.0000
18 1 0.0000
19 0 0.0000
19 1 0.0000
20 0 0.0000
20 1 0.0000

```



22	0	0.0000
22	1	0.0000
24	0	0.0000
24	1	-0.2203
24	2	0.2203
25	0	0.0000
25	1	0.0000
26	0	0.0000
26	1	0.0000
28	0	0.0000
28	1	0.0000
31	0	0.0000
31	1	0.0000
32	0	0.0000
32	1	0.0000
35	0	0.0000
35	1	0.0000
36	0	0.0000
36	1	0.0000
37	0	0.0000
37	1	0.0000
38	0	0.0000
38	1	-0.9905
38	2	0.9905

Appendix E – Winsteps Score File (Grade 3 Reading)

PERSON SCORE FILE FOR D:\Data\Kentucky\KPREP2015\WINSTEP\d03RDv0.con Jul 8 11:54 2015

USCALE=1.00

SCORE	MEASURE	S.E.	INFO	NORMED	S.E.	FREQUENCY	%	CUM.FREQ.	%	PERCENTILE
0	-5.4810	2.0152	.25	-65	181	160	.3	160	.3	1
1	-4.0496	1.0292	.94	63	92	3	.0	163	.3	1
2	-3.2983	.7464	1.79	131	67	6	.0	169	.3	1
3	-2.8365	.6237	2.57	172	56	9	.0	178	.3	1
4	-2.4936	.5520	3.28	203	50	21	.0	199	.4	1
5	-2.2161	.5041	3.94	228	45	48	.1	247	.5	1
6	-1.9799	.4696	4.54	249	42	78	.2	325	.6	1
7	-1.7719	.4435	5.08	268	40	140	.3	465	.9	1
8	-1.5845	.4232	5.58	285	38	245	.5	710	1.4	1
9	-1.4124	.4069	6.04	300	37	375	.7	1085	2.1	2
10	-1.2523	.3938	6.45	314	35	480	.9	1565	3.1	3
11	-1.1015	.3831	6.81	328	34	633	1.2	2198	4.3	4
12	-.9582	.3744	7.14	341	34	712	1.4	2910	5.7	5
13	-.8208	.3672	7.42	353	33	834	1.6	3744	7.3	7
14	-.6882	.3613	7.66	365	32	931	1.8	4675	9.2	8
15	-.5594	.3566	7.86	377	32	1025	2.0	5700	11.2	10
16	-.4335	.3530	8.03	388	32	1131	2.2	6831	13.4	12
17	-.3100	.3502	8.15	399	31	1228	2.4	8059	15.8	15
18	-.1881	.3483	8.24	410	31	1393	2.7	9452	18.5	17
19	-.0672	.3471	8.30	421	31	1519	3.0	10971	21.5	20
20	.0531	.3467	8.32	432	31	1628	3.2	12599	24.7	23
21	.1734	.3470	8.30	442	31	1814	3.6	14413	28.3	26
22	.2942	.3481	8.25	453	31	2014	4.0	16427	32.2	30
23	.4159	.3499	8.17	464	31	2075	4.1	18502	36.3	34
24	.5393	.3526	8.04	475	32	2321	4.6	20823	40.9	39

25	.6647	.3560	7.89	486	32	2481	4.9	23304	45.7	43
26	.7930	.3605	7.70	498	32	2422	4.8	25726	50.5	48
27	.9249	.3660	7.47	510	33	2648	5.2	28374	55.7	53
28	1.0613	.3728	7.20	522	33	2722	5.3	31096	61.0	58
29	1.2032	.3810	6.89	535	34	2721	5.3	33817	66.3	64
30	1.3522	.3911	6.54	548	35	2689	5.3	36506	71.6	69
31	1.5099	.4036	6.14	562	36	2602	5.1	39108	76.7	74
32	1.6788	.4190	5.70	577	38	2512	4.9	41620	81.7	79
33	1.8624	.4385	5.20	594	39	2336	4.6	43956	86.2	84
34	2.0655	.4638	4.65	612	42	2052	4.0	46008	90.3	88
35	2.2957	.4974	4.04	633	45	1760	3.5	47768	93.7	92
36	2.5658	.5446	3.37	657	49	1426	2.8	49194	96.5	95
37	2.8996	.6157	2.64	687	55	974	1.9	50168	98.4	97
38	3.3506	.7384	1.83	727	66	529	1.0	50697	99.5	99
39	4.0886	1.0221	.96	794	92	224	.4	50921	99.9	99
40	5.5084	2.0111	.25	921	180	49	.1	50970	100.0	99

27	R3087	-0.9126	-0.9126	0	0	0.0000	0.0000	0
28	R3089	1.9051	1.9051	0	0	0.0000	0.0000	0
29	R3091	-1.1982	-1.1982	0	0	0.0000	0.0000	0
30	R3250	0.9736	0.9736	0	0	0.0000	0.0000	0
31	R3251	0.5739	0.5739	0	0	0.0000	0.0000	0
32	R3253	-0.6594	-0.6594	0	0	0.0000	0.0000	0
33	R3254	-1.0295	-1.0295	0	0	0.0000	0.0000	0
34	R3255	-0.1466	-0.1466	0	0	0.0000	0.0000	0
35	R3257	0.5111	0.5111	0	0	0.0000	0.0000	0
36	R3258	1.4261	1.4261	0	0	0.0000	0.0000	0
37	R3262	0.2362	0.2362	0	0	0.0000	0.0000	0
38	R3263	0.4516	0.4516	0	0	-0.9905	-0.9905	0.9905	0.9905	0