HumRRO
HUMAN RESOURCES RESEARCH ORGANIZATION

2023 No. 165

Third-Party Checking of 2023 Scaling and Equating for the Alternate Kentucky Summative Assessments (AKSA)

Authors: Gertrudes Velasquez
Meng Fan
Emily Dickinson
Art Thacker

Third-Party Checking of 2023 Scaling and Equating for the Alternate Kentucky Summative Assessments (AKSA)

Table of Contents

Executive Summary

Education Measurement Consulting (Dr. William Auty) and the Human Resources Research Organization (HumRRO) independently calibrated, scaled, and equated the 2023 Alternate Kentucky Summative Assessments (AKSA) and produced the raw-score-to-theta-score tables that were applied to students' test results. Results calculated by HumRRO were identical to those calculated by Dr. Auty. Given that HumRRO's results were identical to Dr. Auty's, we are assured that no processing errors were committed.

Third-Party Checking of 2023 Scaling and Equating for the Alternate Kentucky Summative Assessments (AKSA) Tests

Introduction

The Alternate Kentucky Summative Assessments (AKSA) assesses student performance on the Alternate Assessment Targets. Alternate Assessment Targets represent limits to selected Kentucky Academic Standards (KAS). The AKSA is administered to students with moderate and significant disabilities (i.e., less than 1% of the total student population for whom traditional assessments would not provide a reliable and valid measure of progress). The Kentucky Department of Education (KDE) contracted with the Human Resources Research Organization (HumRRO) to provide independent, external replication of the psychometric processing activities for both the AKSA and the Kentucky Summative Assessment (KSA).[1]

In spring 2023, the AKSA was administered in reading, mathematics, science, social studies, editing and mechanics, and on-demand writing. Reading and mathematics were assessed in Grades 3 through 8 and Grade 10; science in Grades 4, 7, and 11; social studies, editing and mechanics, and on-demand writing in Grades 5, 8, and 11.

This report describes how student test responses for the 2023 AKSA 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 not conducted independently, but cooperatively by both HumRRO and Education Measurement Consulting (Dr. William Auty). Several interim checks were conducted during the analyses, and any discrepancies between the two companies were investigated and ultimately resolved. This process was conducted transparently among the University of Kentucky (the test vendor), Dr. Auty, HumRRO, and the Kentucky Department of Education (KDE) via frequent email communications and conference calls. The process was guided by a specifications document created by Dr. Auty and regularly updated based on decisions before and during calibration. This documentation is vital for ensuring consistency of processing across years and for guiding psychometric processing in future years.

Brief Description of 2023 Assessments

The AKSA consists of picture-based, 3-option multiple-choice assessments that are fully scripted and read by a qualified test administrator. It is administered over two testing windows. For each content area, 15 items are administered in fall, and 15 items are administered in spring.

In the 2022-2023 academic year, the AKSA transitioned from a raw score to a scale score reporting model. Scores are reported on a scale of 150-250. An equating procedure was implemented to ensure that scores were comparable to those from the prior year.

---

[1] The AKSA is composed of two tests. The Attainment Task parallels the KSA, whereas the Transition Attainment Record parallels the ACT. Accordingly, references to the AKSA in this report focus on the Attainment Task component only.

Analysis Procedures

Item parameters were generated (i.e., calibrated) separately for each grade/subject combination. Common item equating was then used to put the NAPD cut points established in 2022 onto this new scale. Finally, raw-score-to-scale (RSS) score tables were generated for all assessments. For each of these analyses, HumRRO followed the analysis specifications provided by Dr. Auty, independently conducted analyses, and verified that the two sets of results matched. Calibration and equating procedures were conducted using the *mirt* (v1.4.1; Chalmers, 2012) package in R (R Core Team, 2022) statistical software. We summarize HumRRO's processes and procedures for conducting these analyses below.

Sample Identification and File Construction

We received the data that had already applied exclusion rules to select the sample of student responses to include in the calibration analyses.[2] Kentucky selects most of its student population for use in the calibration sample for scaling and equating. However, some students are purposefully excluded from the calibration samples, as specified in the 2023 psychometrics specifications document.[3] Kentucky's exemption rules generally only apply to students who receive accommodations (e.g., Braille forms, audio, large print, etc.), students with duplicate records (the same identification number and name), and students with blank total test score values. HumRRO removed students with all missing responses. Dr. Auty and HumRRO verified the n-counts for the calibration sample.

The next step was to format all subject/grade files to be read into R software for processing. An example of a student response file is provided in Appendix A. HumRRO and Dr. Auty used the same raw student data files (containing all student responses) but did not share programming or methodology for creating the scored data files. Prior to item calibration, item responses were scored into correct/incorrect responses using response keys.

Calibration and Scaling Procedures

Once scored data files were prepared, items were calibrated using the Rasch model (Rasch, 1960). This step produced the item difficulty parameter. These parameters are produced on the student ability (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. Dr. Auty and HumRRO verified item parameter estimates after this step.

Equating Procedures

For each assessment, we also placed the NAPD cut points that were set in 2022 onto the 2023 reporting scale. To accomplish this, we first examined the quality of items for inclusion in the equating procedure. These anchor items were items common to the 2022 and 2023 AKSA administrations. Items were flagged for removal from the anchor set if they had negative item slopes (item difficulty parameters, $a < -0.01$) or demonstrated poor item fit (based on Orlando and Thissen's (2000) signed chi-squared test, S_X2) in the 2023 calibration. Items were also removed if they performed well in 2022 but demonstrated parameter drift in 2023. We limited the total items removed to 25% of the original anchor set.

---

[2] Students who are excluded from calibration analyses are not excluded from scoring and reporting.

[3] Kentucky Spring 2023 Psychometric Analysis Specifications v2.1.

The following steps were applied for the equating procedures.

1. Item Calibration: Calibrate the 2022 test items by using the responses provided by students in 2022. One item in reading Grade 4 (item F4) was removed from the 2022 item calibration because of issues related to teacher directions. The item was retained in the 2023 item calibration.

2. Data Exclusion: Remove any cases with missing responses to ensure the accuracy of the calibration process.

3. Theta Value Estimation: Use the EAPsum method to estimate the 2022 theta values, aiming for a single theta value for each raw score.

4. Establish NAPD Cuts: Create NAPD cuts on the 2023 scale to match the raw score cuts observed in 2022.

5. Application of Cuts: Implement these established cuts from 2022 onto the 2023 results to categorize or interpret the 2023 outcomes based on the previously set criteria.

Raw-score-to-Scale-Score Procedures

The item parameter estimates from the 2023 calibration were used to generate person ability estimates. Ability estimates were reported in raw-score-to-theta-score tables.

Once theta scoring tables were obtained, they were linearly transformed to a reporting scale of 150-250 for all grade/subjects. The scale score transformation equation is given by

$$SS = 20 * \theta + 200,$$

where SS represents students' scale scores and θ represents students' ability estimates. The scale score slope and intercept transformation constants are 20 and 200, respectively.

Performance levels (Novice, Apprentice, Proficient, and Distinguished; NAPD) were also assigned to each score, based on the cut points identified via equating procedures.

Verification of 2023 Scoring Tables

After verification of the raw-score-to-scale score tables, scoring tables were generated to assign student performance level classifications. HumRRO checked the 2023 scoring tables and verified that the correct scale score ranges were associated with each performance level. HumRRO matched Dr. Auty on all subjects and grades.

Documentation

As HumRRO and Dr. Auty completed each step of the process described above, item parameters, anchor items, score, and output files were shared to check for inconsistencies. Output files included the number of cases in the calibration sample, item-level information (i.e., item parameters), and the theta scoring tables. A sample of the input and output files is appended to this document. They include:

1. Student Response Data (Appendix A). These files contain the student response data. The appendix includes an example of student response data for Science Grade 4.

2. Item Parameter (Appendix B). These files contain the item parameters for the operational items. The file included in the appendix is an example of an item parameter file for Science Grade 4.

3. Raw-Score-to-Scale Score (Appendix C). The file includes raw-score-to-scale score tables. The file included in the appendix is an example of a raw-score-to-scale score table for Editing and Mechanics Grade 5.

4. Scale Score and Performance Level (Appendix D). The file contains the scale score ranges and their corresponding performance levels.

Conclusion

Education Measurement Consulting (Dr. William Auty) and HumRRO independently calculated the scaled/equated raw-score-to-scale-score tables for 2023 AKSA reading and math (Grades 3-8, 10), social studies and writing (Grades 5, 8, 11), and science (Grades 4, 7, 10). No differences were found between the two sets of parameter estimations or raw-score-to-scale-score tables. Given that HumRRO's and Dr. Auty's scaling and equating results were identical, HumRRO is confident that no processing errors were committed.

References

Chalmers, R. P. (2012). Mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software, 48*(6), 1-29. https://doi.org/10.18637/jss.v048.i06.

Orlando, M. &Thissen, D. (2000). Likelihood-based item fit indices for dichotomous item response theory models. *Applied Psychological Measurement*, 24, 50-64.

R Core Team (2023). R: A language and environment for statistical computing. R Foundatiaon for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Chicago: University of Chicago Press.

Appendix A – Student Response Data (Science Grade 4)

GRADE	SSID	Science_A1	Science_A2	Science_A3	Science_A4	Science_A5	Science_B1	Science_B2	Science_B3	Science_B4	Science_B5
4	2121006961	a	b	c	c	b	a	c	c	c	c
4	2120626441	a	c	c	a	a	a	b	b	b	a
4	2120613553	c	a	b	c	c	a	c	c	a	a
4	2120543974	b	b	a	a	b	b	a	a	b	b
4	2120559978	a	a	a	a	a	b	b	b	b	b
4	2120563987	c	c	a	c	c	b	a	b	c	c
4	2120622320	c	a	a	a		c	c	a	b	a
4	2120667997	c	a	b	b	b	c	c	a	c	c
4	2120696199	b	b	b	b	b	b	b	b	b	b
4	2120749458	c	b	a	c	b	c	b	a	a	c
4	2120630336	a	a	a	a	a	a	a	a	a	a
4	2120607947	a	c	b	b	a	a	b	a	a	a
4	2120754968	b	b	b	b	b	b	b	b	b	b
4	2120626794	a	a	c	c	c	c	c	c	c	c
4	2120570737	a	c	c	c			c	c	a	c
4	2120599867	c	c	c	c	c	c	c	c	c	c
4	2120626977	c	b	b	b	b	c	a	a	b	c
4	2120632412	a	c	b	a	a	b	b	a	a	a
4	2120594437	b	c	b	b	c	c	b	b	c	b
4	2120725791	a	a	a	a	b	a	c	b	b	b
4	2120961916	a	c	b	a	c	c	b	c	c	a
4	2120829182	c	c	c	c	c	a	c	b	c	c
4	2120573691	a	c	b	c	a	b	b	c	b	a
4	2120797586	b	c	b	b	b	b	c	a	c	b
4	2120756939	c	a	b	c	b	c	c	c	b	a
4	2120595135	c	c	c	c	c	c	c	c	c	c
4	2120532587	c	c	b	c	c	b	b	a	b	c
4	2120618566	c	c	b	a	a	a	b	a	c	c

Appendix B – Item Parameters (Science Grade 4)

item	class	name	parnum	h_value	parm_diff
A1	dich	a1	1	1.0000	0
A1	dich	d	2	-0.3637	0
A1	dich	g	3	0.0000	0
A1	dich	u	4	1.0000	0
A2	dich	a1	5	1.0000	0
A2	dich	d	6	-0.0062	0
A2	dich	g	7	0.0000	0
A2	dich	u	8	1.0000	0
A3	dich	a1	9	1.0000	0
A3	dich	d	10	0.0275	0
A3	dich	g	11	0.0000	0
A3	dich	u	12	1.0000	0
A4	dich	a1	13	1.0000	0
A4	dich	d	14	-0.1666	0
A4	dich	g	15	0.0000	0
A4	dich	u	16	1.0000	0
A5	dich	a1	17	1.0000	0
A5	dich	d	18	-0.4335	0
A5	dich	g	19	0.0000	0
A5	dich	u	20	1.0000	0
B1	dich	a1	21	1.0000	0
B1	dich	d	22	-0.1159	0
B1	dich	g	23	0.0000	0
B1	dich	u	24	1.0000	0
B2	dich	a1	25	1.0000	0
B2	dich	d	26	-0.3463	0
B2	dich	g	27	0.0000	0
B2	dich	u	28	1.0000	0
B3	dich	a1	29	1.0000	0
B3	dich	d	30	0.0022	0
B3	dich	g	31	0.0000	0
B3	dich	u	32	1.0000	0
B4	dich	a1	33	1.0000	0
B4	dich	d	34	0.0527	0
B4	dich	g	35	0.0000	0
B4	dich	u	36	1.0000	0
B5	dich	a1	37	1.0000	0
B5	dich	d	38	-0.5762	0
B5	dich	g	39	0.0000	0
B5	dich	u	40	1.0000	0
C1	dich	a1	41	1.0000	0
C1	dich	d	42	-0.6866	0
C1	dich	g	43	0.0000	0
C1	dich	u	44	1.0000	0
C2	dich	a1	45	1.0000	0
C2	dich	d	46	0.0612	0
C2	dich	g	47	0.0000	0
C2	dich	u	48	1.0000	0
C3	dich	a1	49	1.0000	0
C3	dich	d	50	-0.3898	0
C3	dich	g	51	0.0000	0
C3	dich	u	52	1.0000	0
C4	dich	a1	53	1.0000	0
C4	dich	d	54	0.0527	0
C4	dich	g	55	0.0000	0
C4	dich	u	56	1.0000	0
C5	dich	a1	57	1.0000	0
C5	dich	d	58	-0.2006	0
C5	dich	g	59	0.0000	0
C5	dich	u	60	1.0000	0
D1	dich	a1	61	1.0000	0
D1	dich	d	62	-0.3723	0
D1	dich	g	63	0.0000	0

```
D1      dich    u       64      1.0000  0
D2      dich    a1      65      1.0000  0
D2      dich    d       66      -0.4072 0
D2      dich    g       67      0.0000  0
D2      dich    u       68      1.0000  0
D3      dich    a1      69      1.0000  0
D3      dich    d       70      -0.8992 0
D3      dich    g       71      0.0000  0
D3      dich    u       72      1.0000  0
D4      dich    a1      73      1.0000  0
D4      dich    d       74      -0.0062 0
D4      dich    g       75      0.0000  0
D4      dich    u       76      1.0000  0
D5      dich    a1      77      1.0000  0
D5      dich    d       78      -0.6310 0
D5      dich    g       79      0.0000  0
D5      dich    u       80      1.0000  0
E1      dich    a1      81      1.0000  0
E1      dich    d       82      -0.0568 0
E1      dich    g       83      0.0000  0
E1      dich    u       84      1.0000  0
E2      dich    a1      85      1.0000  0
E2      dich    d       86      -0.6494 0
E2      dich    g       87      0.0000  0
E2      dich    u       88      1.0000  0
E3      dich    a1      89      1.0000  0
E3      dich    d       90      -1.2557 0
E3      dich    g       91      0.0000  0
E3      dich    u       92      1.0000  0
E4      dich    a1      93      1.0000  0
E4      dich    d       94      0.3159  0
E4      dich    g       95      0.0000  0
E4      dich    u       96      1.0000  0
E5      dich    a1      97      1.0000  0
E5      dich    d       98      -0.5853 0
E5      dich    g       99      0.0000  0
E5      dich    u       100     1.0000  0
F1      dich    a1      101     1.0000  0
F1      dich    d       102     0.3073  0
F1      dich    g       103     0.0000  0
F1      dich    u       104     1.0000  0
F2      dich    a1      105     1.0000  0
F2      dich    d       106     -0.4072 0
F2      dich    g       107     0.0000  0
F2      dich    u       108     1.0000  0
F3      dich    a1      109     1.0000  0
F3      dich    d       110     0.4638  0
F3      dich    g       111     0.0000  0
F3      dich    u       112     1.0000  0
F4      dich    a1      113     1.0000  0
F4      dich    d       114     0.1456  0
F4      dich    g       115     0.0000  0
F4      dich    u       116     1.0000  0
F5      dich    a1      117     1.0000  0
F5      dich    d       118     -0.8104 0
F5      dich    g       119     0.0000  0
F5      dich    u       120     1.0000  0
```

Appendix C – Raw-Score-to-Scale Score (Science Grade 4)

test	raw	theta	thetaSE	ScaleScore	SE	PL
SC_04	0	-1.21035	0.325364	188	3.25	N
SC_04	1	-1.10598	0.320803	189	3.21	N
SC_04	2	-1.00444	0.316559	190	3.17	N
SC_04	3	-0.90549	0.312633	191	3.13	N
SC_04	4	-0.80889	0.309024	192	3.09	N
SC_04	5	-0.71442	0.305728	193	3.06	N
SC_04	6	-0.62188	0.30274	194	3.03	N
SC_04	7	-0.53105	0.300055	195	3	N
SC_04	8	-0.44175	0.297667	196	2.98	N
SC_04	9	-0.35378	0.29557	196	2.96	N
SC_04	10	-0.26697	0.293758	197	2.94	N
SC_04	11	-0.18114	0.292225	198	2.92	A
SC_04	12	-0.09612	0.290967	199	2.91	A
SC_04	13	-0.01176	0.289979	200	2.9	A
SC_04	14	0.072106	0.289258	201	2.89	A
SC_04	15	0.155632	0.288802	202	2.89	A
SC_04	16	0.23897	0.288609	202	2.89	A
SC_04	17	0.322272	0.288677	203	2.89	P
SC_04	18	0.40569	0.289008	204	2.89	P
SC_04	19	0.489374	0.289601	205	2.9	P
SC_04	20	0.573478	0.290458	206	2.9	P
SC_04	21	0.658158	0.291583	207	2.92	P
SC_04	22	0.743573	0.292978	207	2.93	P
SC_04	23	0.829886	0.294648	208	2.95	P
SC_04	24	0.917265	0.296597	209	2.97	P
SC_04	25	1.005884	0.298831	210	2.99	D
SC_04	26	1.095926	0.301355	211	3.01	D
SC_04	27	1.187578	0.304175	212	3.04	D
SC_04	28	1.281038	0.307297	213	3.07	D
SC_04	29	1.376511	0.310725	214	3.11	D
SC_04	30	1.474211	0.314464	215	3.14	D

Appendix D – Scale Score and Performance Level

Test	h_Min_N	h_Max_N	h_Min_A	h_Max_A	h_Min_P	h_Max_P	h_Min_D	h_Max_D
EM 05	150	197	198	204	205	208	209	250
EM 08	150	192	193	206	207	221	222	250
EM 11	150	192	193	205	206	218	219	250
MA 03	150	191	192	205	206	220	221	250
MA 04	150	192	193	204	205	218	219	250
MA 05	150	191	192	206	207	221	222	250
MA 06	150	191	192	204	205	216	217	250
MA 07	150	193	194	202	203	212	213	250
MA 08	150	196	197	201	202	209	210	250
MA 10	150	192	193	202	203	214	215	250
RD 03	150	191	192	206	207	221	222	250
RD 04	150	190	191	205	206	223	224	250
RD 05	150	190	191	204	205	221	222	250
RD 06	150	181	182	203	204	228	229	250
RD 07	150	189	190	204	205	224	225	250
RD 08	150	187	188	205	206	226	227	250
RD 10	150	190	191	207	208	226	227	250
SC 04	150	193	194	204	205	216	217	250
SC 07	150	189	190	202	203	217	218	250
SC 11	150	192	193	206	207	221	222	250
SS 05	150	194	195	204	205	218	219	250
SS 08	150	189	190	204	205	219	220	250
SS 11	150	188	189	205	206	220	221	250
WR 05	150	195	196	203	204	208	209	250
WR 08	150	194	195	204	205	215	216	250
WR 11	150	192	193	204	205	215	216	250