jMetrik Step-by-step

Trevor A. Holster

This is a step-by-step guide on how to import data into jMetrik and conduct some simple analyses. You should purchase the manual (Meyer, 2014) if you want to do more advanced analyses.

1. Download and Install jMetrik

jMetrik is a free application that runs on any Windows, MacOS, or Linux computer **that has Java installed**. jMetrik is available from this url:
https://itemanalysis.com/jmetrik-download/

Make sure that you updated Java or you may have problems getting jMetrik to run properly. You can download Java from this URL:
https://www.java.com/en/download/

The datafile that we will use in this workshop is available from this url:
https://tinyurl.com/yb4ybcs6

The data that we will analyze came from a pilot administration of a proposed placement test. You can download a sample test here:
https://tinyurl.com/32fjm6mc

There are accompanying videos on YouTube:
https://tinyurl.com/ynt73bw3

2. Create a Database and Import Data

Click **Manage >>> New Database.** Name your database using lowercase letters. I will name my new database "jaltdemo." Open the database by clicking **Manage >>> Open Database**.

Click **Manage >>> Import Data**, then browse to the sample database and select the jMetrik Data.csv file as the data file. Give it a table name. I will call mine jaltdemo.

Click **"Import"** to import the data. This data is now saved in the new database and will be there the next time you open this database in jMetrik.

Data Layout

The first column is student codes, running from S1201 to S5334. There are 281 students.

The top row is item codes. The test has four sections, with 130 items:

1. Items 1-16: Coded LC (listening cloze) 16 items Response codes A to T
2. Items 17-36: Coded LD (listening dictation) 20 items Response codes A to T
3. Items 37-76: Coded CE (cloze elide) 40 items Response codes 0 and 1
4. Items 77-130: Coded V (vocabulary) 54 items Response codes A to E

Missing data

Missing data should be coded NA. The sample dataset does not contain missing data.

3. Creating an Answer Key Using Basic Item Scoring

Click **Transform >> Basic Item Scoring**. The top row shows the item codes. The answer key for each item needs to be entered into the second row and the number of response options needs to be entered into the third row. We just need to copy the row of answer keys from Table 1 and 2 into the second row of the wizard and enter 24 in the third row for every item.

These two listening sections use response codes of A to T, plus X for skipped items. In this case, we want to score X as an incorrect response, so the number of response options should be 24, to include all the letters from A to X.

The Tab key will let you move to the next box in the row. You can use CTL * C to copy the contents of a cell and CTL +V to paste it into another cell. Once all the data entry for sections 1 and 2 is complete, click OK to save the answer key.

Table 1: Listening Cloze Answer Key

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Code	LC	LC	LC	LC	LC	LC	LC	LC	LC	LC	LC	LC	LC	LC	LC	LC
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Key	O	G	B	C	J	L	P	S	T	M	K	F	R	I	N	D

Table 2: Listening Dictation Answer Key

Item	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
Code	LD	LD	LD	LD	LD	LD	LD	LD	LD	LD	LD	LD	LD	LD	LD	LD	LD	LD	LD	LD
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Key	I	H	L	J	Q	C	D	N	O	P	B	R	M	G	S	F	K	E	T	A

4. Creating an Answer Key Using Advanced Item Scoring

Items with the same answer key can be processed together using the Advanced Item Scoring wizard. Table 3 lists all the vocabulary items with the answer key A, so we can enter all these together. This is much faster than entering each one separately.

Table 3: Vocabulary Answer Key A

Item	84	91	93	94	100	109	114	115	116	119	122	123	124	125	127
Code	V8	V15	V17	V18	V24	V33	V38	V39	V40	V43	V46	V47	V48	V49	V51
Key	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A

Click **Transform >> Advanced Item Scoring.** First, enter all the characters from A to E, plus X in the "Option" column, In the "Score" column, enter a score of 1 for the option A, and a score of 0 for all the other options. Select item V8 and click the > button to move it to the selection panel. On a Windows computer, you can hold down the CTRL key to select multiple items, so you can select all the items listed above and move them to the selection panel. On a Mac, you use the Command button for this. Finally, click "Submit'. You will see the selected items highlighted in bold and the scoring syntax at the bottom. Then click OK to save the answer key for these items.

Now we need to repeat this for all the items with "B" as the scoring key (Table 4), then "C" (Table 5), "D" (Table 6), and finally "E" (Table 7), with the correct response scored as 1 and the other responses scored as 0.

Table 4: Vocabulary Answer Key B

Item	82	89	96	104	105	107	118	130
Code	V6	V13	V20	V28	V29	V31	V42	V54
Key	B	B	B	B	B	B	B	B

Table 5: Vocabulary Answer Key C

Item	78	79	81	83	86	87	92	110	120	126	128
Code	V2	V3	V5	V7	V10	V11	V16	V34	V44	V50	V52
Key	C	C	C	C	C	C	C	C	C	C	C

Table 6: Vocabulary Answer Key D

Item	77	80	85	88	95	99	101	102	103	106	111	112	121	129
Code	V1	V4	V9	V12	V19	V23	V25	V26	V27	V30	V35	V36	V45	V53
Key	D	D	D	D	D	D	D	D	D	D	D	D	D	D

Table 7: Vocabulary Answer Key E

Item	90	97	98	108	113	117
Code	V14	V21	V22	V32	V37	V41
Key	E	E	E	E	E	E

5. Scoring Numerically Coded Rating Scale Items

Section 3 of the test uses a format called *cloze elide* (CE). These use a numeric key, not an alphabetic one. Some items (*unplanned items*) are reverse scored, they are listed in Table 8.

Table 8: Cloze Elide Unplanned Items Answer Key

Item	37	38	39	42	43	48	49	52	54	55	56	60	62	64	65	68	69	70	73	76
Code	CE	CE	CE	CE	CE	CE	CE	CE	CE	CE	CE	CE	CE	CE	CE	CE	CE	CE	CE	CE
	1	2	3	6	7	12	13	16	18	19	20	24	26	28	29	32	33	34	37	40
Key	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Open the Advanced Item Scoring wizard and enter “0” and “1” as the response options, with scores of 1 and 0 respectively (i.e. response option “0” has a score of 1; response option “1” has a score of 0).

The *planned items* use a more regular scoring key, with a response of “1” scored as 1 and a response of “0” scored as zero. These are listed in Table 9. Use the Advanced Scoring wizard to score these, with a response of “0” having a score of 0 and a response of “1” having a score of 1.

Table 9: Cloze Elide Planned Items Answer Key

Item	40	41	44	45	46	47	50	51	53	57	58	59	61	63	66	67	71	72	74	75
Code	CE	CE	CE	CE	CE	CE	CE	CE	CE	CE	CE	CE	CE	CE	CE	CE	CE	CE	CE	CE
	4	5	8	9	10	11	14	15	17	21	22	23	25	27	30	31	35	36	38	39
Key	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Scoring Polytomous Items

The sample data only contains dichotomous items, with scores of 0 or 1. Items with polytomous scales can be scored using the advanced item scoring wizard in much the same way as these dichotomous items.

6. Generating Student Scores

Click **Transform >>> Test Scaling.** In the *Score* section, select *Sum Score* as the score type. Give this a name, for example "total". Select the items you want to include in the test score. You can select all items on a Windows computer using CTRL + A or you can select a range of items by holding down the SHIFT key when you click on the first and last item. Click the *Run* button. In the Data tab, click *Refresh* at the top. There is a new variable named "total" that shows the total number correct for each student.

To calculate the percentage scores, select *Average Score* from the *Transform* menu. I'll call this "percentage". Select all the items and click "Run". Refresh the data display.

Exporting and saving results

Click **Manage >>> Export Data.** Make sure that "Comma" is selected from the Delimiter panel. Select CSV Files as the file type and give the file a meaningful name, "Test Scores", for example. This will save a new data file with the students scores included.

7. Test Reliability

Click **Analyze >>> Item Analysis**. Select all the items and select *Polyserial correlation* in the Item-total Correlation type. Click *Run* and the results will appear in a new tab. You can save this as a text file from the File menu by clicking "Save" or "Save As".

Test Level Statistics are provided at the bottom of the output, shown in Table 10. This includes descriptive statistics and a *Reliability Analysis* section. The coefficient alpha of .91 is excellent for a classroom test. As a rule-of-thumb, values below .80 would generally be considered low, while high-stakes decisions would typically require values of .90 or higher.

Table 10

```
TEST LEVEL STATISTICS
=======================================
Number of Items = 130
Number of Examinees =        281
Min = 38.0000
Max = 114.0000
Mean = 78.9929
Median = 80.0000
Standard Deviation = 13.9817
Interquartile Range = 18.0000
Skewness = -0.0427
Kurtosis = -0.2297
KR21 = 0.8480
=======================================
```

Table 11

RELIABILITY ANALYSIS

Method	Estimate	95% Conf. Int.	SEM
Guttman's L2	0.9120	(0.8967, 0.9260)	4.1558
Coefficient Alpha	0.9062	(0.8900, 0.9212)	4.2888
Feldt-Gilmer	0.9093	(0.8937, 0.9238)	4.2172
Feldt-Brennan	0.9091	(0.8933, 0.9236)	4.2233
Raju's Beta	0.9062	(0.8900, 0.9212)	4.2888

8. Classical Item Analysis

The *Item Analysis* output can be used to identify poorly functioning items. Table 10 shows statistics for four items, LC1, LC2, CE12, and V42. The *Discrimin* column shows the item discrimination. We would usually want items with discriminations of .40 or higher. The *Difficulty* column in Table 12 shows the proportion of correct responses, in this case, 0.8968. In this classical analysis, a higher value for difficulty indicates an easier item because more students answered correctly.

Item LC1 has a low discrimination of .15, meaning that it does a poor job of discriminating between high and low-proficiency students. In the case of LC1, the item is very easy. Only 10% of students did not answer correctly, so the reason for the low discrimination is probably that the item is too easy for this sample of students. However, it does have a positive discrimination and is the first item in the test, so retaining it would be justified on the grounds that it provides a confidence boost to begin the test.

Item LC2 has a reasonably good discrimination of .54 and difficulty of 0.72, so this item is functioning acceptably, but is fairly easy for these students. Item CE12 has a slightly negative discrimination of -.04, and is also very easy. It is a cloze elide shadowing item, but we can see that it is an *unplanned item*. The unplanned CE items are necessary for the functioning of this test format because they function as distractors, but they are very poor at discriminating between high and low-ability students.

Item V42 also has a discrimination very close to zero and is a very difficult item, with only 14% of students answering correctly. It is a 5-option multiple-choice question so we would expect 20% of students to answer correctly through random guessing. This indicates a malfunctioning distractor. the response option E, a distractor, attracted 49% of responses. Even more concerning, the distractor E had a discrimination of .15 compared to .03 for the item key, option B. The problem with this item is related to the distractor E, *Highlight*. It is possible that this item would function well with a sample of students of higher proficiency, but this distractor should be replaced.

Item V42: Look
A) Echo
B) Peer
C) Rent
D) Exclude
E) Highlight

Table 12
Sample Item Statistics

Item	Option	Score	Difficulty	Std. Dev.	Discrimin.
lc1	Overall		0.8968	0.3048	**0.1517**
	C	0.0	0.0036	0.0597	-0.4529
	E	0.0	0.0036	0.0597	0.0480
	G	0.0	0.0036	0.0597	-0.0952
	I	0.0	0.0107	0.1030	-0.4952
	N	0.0	0.0036	0.0597	-0.5483
	O	**1.0**	**0.8968**	**0.3048**	**0.1517**
	Q	0.0	0.0676	0.2515	-0.0440
	X	0.0	0.0107	0.1030	-0.3511
lc2	Overall		0.7224	0.4486	0.5442
	A	0.0	0.0107	0.1030	-0.4051
	B	0.0	0.0071	0.0842	-0.1026
	C	0.0	0.0391	0.1943	-0.3644
	D	0.0	0.0178	0.1324	-0.3994
	E	0.0	0.0142	0.1187	-0.0275
	F	0.0	0.0107	0.1030	-0.3331
	G	**1.0**	**0.7224**	**0.4486**	**0.5442**
	H	0.0	0.0427	0.2026	-0.3895
	I	0.0	0.0071	0.0842	-0.5912
	J	0.0	0.0071	0.0842	-0.3470
	K	0.0	0.0071	0.0842	-0.1155
	Q	0.0	0.0285	0.1666	-0.0965
	R	0.0	0.0107	0.1030	-0.3871
	S	0.0	0.0071	0.0842	-0.3598
	T	0.0	0.0036	0.0597	-0.0713
	X	0.0	0.0641	0.2453	-0.4746
ce12	Overall		0.8968	0.3048	**-0.0428**
	0.0	**1.0**	**0.8968**	**0.3048**	**-0.0428**
	1.0	0.0	0.1032	0.3048	-0.0309
v42	Overall		**0.1388**	0.3463	**0.0250**
	A	0.0	0.1708	0.3770	-0.0055
	B	**1.0**	**0.1388**	**0.3463**	**0.0250**
	C	0.0	0.0605	0.2388	-0.5323
	D	0.0	0.1246	0.3308	-0.2168
	E	0.0	**0.4947**	0.5009	**0.1537**
	X	0.0	0.0107	0.1030	-0.1979

9. Rasch Analysis

Rasch analysis converts percentage scores into log-odds units, or *logits,* which provide equal interval measures. This is desirable for researchers and also allows person ability and item difficulty to be mapped onto the same measurement scale. The standard introductory text to Rasch analysis is *Applying the Rasch Model* (Bond & Fox, 2015), you are advised to refer to that for more detailed technical explanations.

Conducting Rasch Analysis

Click **Analyze >>> Rasch Models**. From the Global tab, select all the items. The *Center on items* box is checked by default and the *Linear Transformation* section has defaults of Mean = 0, Scale = 1, and Precision = 4. Leave all of these on the default setting. In the Item tab, check *Save item estimates* and enter a meaningful name for the output table. I will call mine *Rasch Items.* In the Person tab, check *Save person fit statistics* and *Save person estimates*. Then click *Run*.

Rasch Item Analysis

Rasch item analysis focuses on *fit statistics* rather than just item correlations and item difficulty is given in logits. A more detailed explanation is given on the next page.

Item difficulty:

Mean item difficulty = 0.00

Useful range = -3 (very easy) to 3 logits (very difficult)

Mean-square fit statistics:

Mean value = 1.00 (approximately)

Misfit = >1.50

WMS = weighted mean-square (infit) – items well-matched to person ability

UMS = unweighted mean-square (outfit) – reflects outlying responses

Table 13 shows the fit statistics for the four items we looked at earlier. Item LC1 is extremely easy, with a logit value of -1.68. The WMS value of 1.05 shows good infit, but the UMS value of 1.54 shows a level of outfit that is of concern. This is probably because the item is very easy and some high-ability students answered incorrectly, perhaps just a single student. Item LC2 is moderately easy and slightly over-fitting. This is consistent with what we saw in the classical item analysis.

Table 13

FINAL JMLE ITEM STATISTICS

Item	Difficulty	Std. Error	WMS	Std. WMS	UMS	Std. UMS
lc1	-1.68	0.20	1.05	0.38	**1.54**	2.23
lc2	-0.37	0.14	0.91	-1.35	0.83	-1.83
ce12	-1.68	0.20	1.10	0.71	**1.95**	3.53
v42	2.75	0.18	1.15	1.27	1.45	2.33

Item CE12 is also very easy and very badly misfitting, with a UMS value of 1.95. This is an unplanned cloze elide item and nearly all students succeeded on this this item. The very high UMS value will be because some high-ability students gave incorrect responses.

Item V42 is extremely difficult, with a logit value of 2.75 and is also moderately misfitting, with a UMS value of 1.45. This item had a badly functioning distractor and very few students answered correctly. The misfit is probably the result of high-ability students being confused by the bad distractor, but low-ability students succeeding through lucky guessing.

Discrimination and Data-Model Fit in Rasch Analysis

Discrimination in Rasch analysis does not mean item correlations as in classical analysis. Discrimination means the slope of the *item characteristic curve* (ICC). The Rasch model is based on the assumption that all items have equal discrimination. Items with higher discrimination will *overfit* the model, items with lower discrimination will *underfit*, or misfit the model. Items with negative correlations will usually badly misfit, but higher item correlations do not automatically mean better fit to the Rasch model. Negative correlations will always indicate a problem, but low positive correlations are not a problem as long as the fit statistics are acceptable.

The mean-square fit statistics have an expected value of 1.00, which indicates well-fitting responses. Values larger than 1.50 indicate a level of misfit that is of concern and values above 2.00 indicate serious problems. Values below 1.00 indicate *overfit*, which means that the item is more consistent than expected.

The *weighted mean-square* (WMS) statistic, or *infit*, is weighted to exaggerate responses where the person ability was well-matched to the item difficulty. These responses provide more information than responses where items are much higher or lower than person ability. A high infit value usually indicates a serious problem with the item.

The *unweighted mean-square* (UMS)statistic, or *outfit*, is not information weighted. High outfit values reflect outlying responses, where high-ability students fail on easy items or low-ability students succeed on difficult items. A common cause of high outfit is that items are extremely easy or extremely difficult, so a very small number of responses can cause extreme items to misfit.

jMetrik also provides standardized fit statistics, which show whether the results are statistically significant, with values outside the range of -2.0 to 2.0 indicating statistical significance($p < .05$). Statistical significance is largely a result of the sample size, so if we have a lot of students, nearly all item misfit will be statistically significant, even if it is not substantively large. The mean-square statistic is therefore much more useful because it shows the substantive size of the misfit, but it is not the only consideration.

Rasch Person Diagnosis

On the data screen in jMetrik, click *Refresh Data View.* You will now see extra columns of data giving the Rasch analysis results for each student. Click **Manage >>> Export Data** to export this data as a file that can be open in Excel.

Person Ability Scores

Person ability is called *theta* in Rasch analysis and is reported in logits, not percentage scores. This makes it easy to compare student ability and item difficulty because they are both given in logits. In the data window, we can see a variable called *sum,* another one called *vsum*, and then *theta.* The *sum* score is the total number correct. We can see that the *vsum* score is always 3 less than the *sum* score. This is because there are three items that every student succeeded on in this test. This is shown in the Rasch output for items CE6, CE40, and V1, where these items are flagged as *Minimum.* Items (or persons) with extreme scores (i.e. 0% or 100%) do not provide any information for the analysis so they are not used in the estimation of Rasch logits.

Rasch Person Fit Statistics

Rasch analysis provides WMS (infit) and UMS (outfit) statistics for persons as well as for items. We can use these for student diagnostics. Table 14 shows the most misfitting students. Four of these students have very high scores and one has a very low score of 38. Student S4426 has a very high UMS (outfit) value of 3.41. This is a very high-ability student, with a total of 114 and a logit ability of 2.98. This high-ability student has failed on some easy items so we would look at their test to see why. For example, they may have been confused by the cloze elide section and made careless errors on that section.

Table 14
Rasch Person Statistics

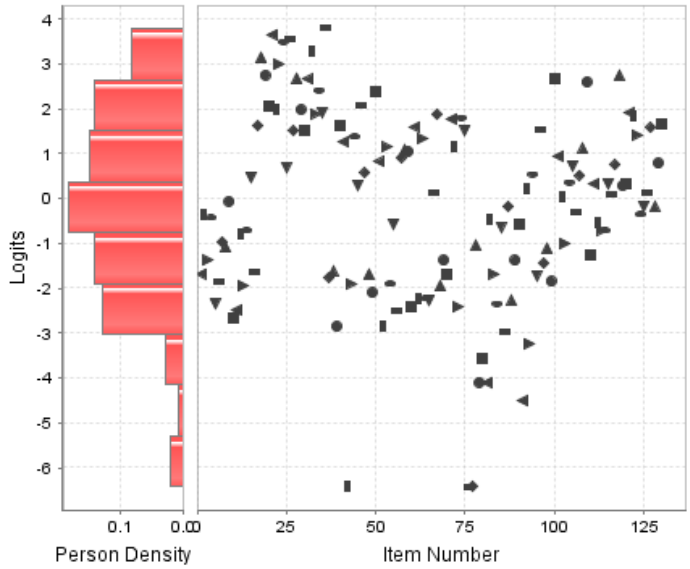
St_No	Total	Sum	Vsum	Theta	Stderr	Extreme	Wms	Stdwms	Ums	Stdums
S4426	114	114	111	**2.98**	0.31	No	1.07	0.43	**3.41**	2.04
S2304	111	111	108	2.72	0.29	No	0.89	-0.64	**2.60**	1.74
S4409	108	108	105	2.48	0.28	No	1.14	0.95	**2.43**	1.77
S4436	107	107	104	2.40	0.27	No	1.14	1.00	**2.27**	1.68
S2403	38	38	35	**-1.58**	0.25	No	1.35	2.51	**2.27**	2.53
S2110	78	78	75	0.64	0.23	No	1.05	0.44	**2.04**	3.03

Student S2403 is a very low-ability student, with a total of 38 and logit ability of -1.58. This student has a very high outfit mean-square (UMS) value of 2.27. This student has succeeded on some difficult items so we would look at their test to try to understand why. For example, they may have struggled with the listening sections but performed much better on the vocabulary section. We would need to look at their test to confirm this, but that is the type of pattern we would look for in diagnosing misfitting students.

Creating a Person-Item Map

Go to Raschitems datascreen. Click **Graph >>> Item Map.** Select *bparam* and in the *Item Parameter Table* box select *Rasch Items.* Click **Run.** jMetrik will produce a person-item map. This compares student ability to item difficulty on the same vertical logit scale, as shown in Figure 1. Mean item difficulty is set to 0.00 logits so a student with ability of 0.00 logits has a 50% chance of succeeding on an item of mean difficulty. The range of item difficulty is quite well matched to the range of student difficulty overall. However, there are very few items below -3 logits so the test cannot measure the very lowest ability students.

Figure 1.
Person-item map. The vertical scale shows the logit scale of item difficulty and person ability.



Comparing the Test Sections

Cloze listening: This section was relatively easy, so it is suitable for separating very low ability students who need remedial instruction from mainstream students. Revised test specifications are needed to address this.

Cloze dictation: The cloze dictation section was much more difficult, so this section is suitable for identifying students who would benefit from more challenging extension classes.

Cloze elide shadowing: This section has a large gap between the relatively difficult planned items and the relatively easy unplanned items. This format is confusing for low-ability students so it is only suitable for higher ability students.

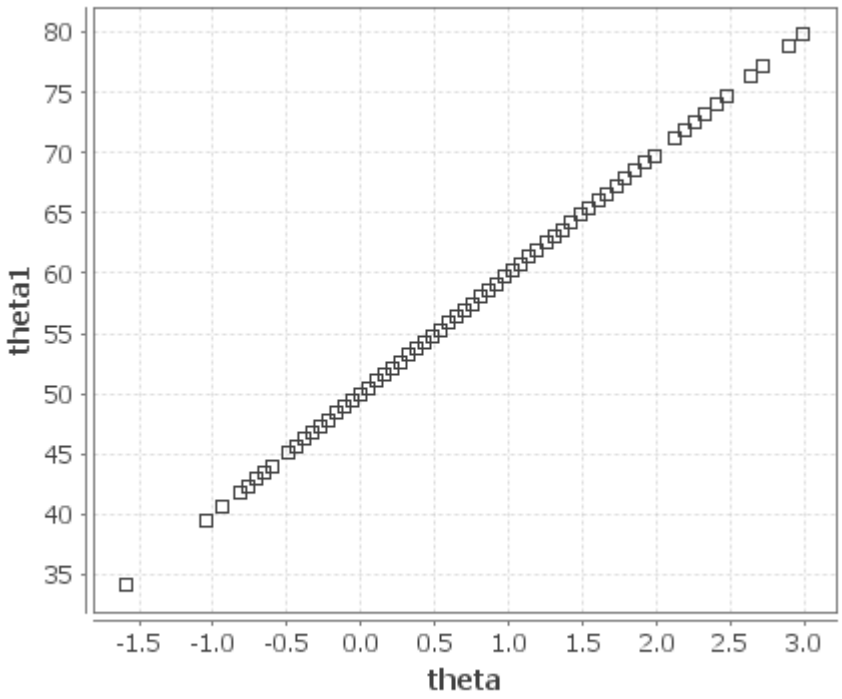
Vocabulary synonymy: These items span a very large range of difficulty, with many easy items near the start of the section and more difficult items towards the end. It would be desirable to replace some medium and high difficulty items with extremely easy items to target remedial students.

Rescaling Rasch Logit Scores

Logit measures of person ability are confusing to most people. We can rescale them to more convenient units. Click **Analyze >>> Rasch Models** and select all the items for analysis . In the *Linear Transformation* section, set *Mean* to 50 and *Scale* to 10. In the *Person* tab, check the bos for *Save person estimates.* This will rescale the logit scores to have a mean item difficulty of 50, with 1 logit rescaled to 10 scaled units. In the data screen, click *Refresh Data View.* There will now be a new theta column that shows student ability on a scale that resembles percentage scores.

Click **Graph >>> Scatterplot.** Select theta (the original logit scores) for the X axis and theta1 (the rescaled scores) for the Y axis, then click *Run.* The scatterplot, shown in Figure 2, shows a perfectly linear transformation of the scores.

Figure 2.
Rescaled logit scores. The logit scores have been rescaled to a more user-friendly scale with mean item difficulty of 50 and 1 logit equal to 10 scaled units.



Rasch Reliability

In the output from theRasch analysis, there is a table called *Scale Quality Statistics*. This includes a reliability coefficient for both items and persons, plus separation and strata indices. The person reliability of .91 is analogous to the Cronbach alpha statistic. The separation index is calculated from the reliability coefficient. The figure of 3.16 means that we can be confident that there are three statistically distinct levels of person ability. What this means is that we have very high confidence that the highest students are actually more proficient than the average students, and also that the average students are actually more proficient than the lowest ability students. In the case of this test, we could confidently use it as a placement test to separate students into two or three different course levels.

Table 15

SCALE QUALITY STATISTICS

Statistic	Items	Persons
Observed Variance	3.5695	0.6242
Observed Std. Dev.	1.8893	0.7901
Mean Square Error	0.0433	0.0568
Root MSE	0.2080	0.2383
Adjusted Variance	3.5263	0.5674
Adjusted Std. Dev.	1.8778	0.7533
Separation Index	9.0278	3.1607
Number of Strata	12.3705	4.5476
Reliability	0.9879	0.9090

Item Reliability

Rasch analysis also provides an item reliability statistic, plus item separation and strata indices. In this case, the item reliability is .98, with a separation index of 9.03. We have very high confidence that the most difficult items are actually more difficult than the easiest items. This test was intended to include some very easy items and some much more difficult items so the item reliability suggests that this was achieved.

References

Aryadoust, V., Ng, L. Y., & Sayama, H. (2021). A comprehensive review of Rasch measurement in language assessment: Recommendations and guidelines for research. *Language Testing*, *38*(1), 6-40. https://doi.org/10.1177/0265532220927487

Bond, T., & Fox, C. (2015). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences, Third Edition*. Taylor & Francis. https://books.google.co.jp/books?id=i13LCQAAQBAJ

Brown, J. D. (2016). *Statistics corner*. JALT Testing and Evaluation Special Interest Group.

Holster, T. A. (2017). Cloze-elide as a classroom reading test. *Shiken*, *21*(2), 1-19. http://teval.jalt.org/sites/teval.jalt.org/files/21_02_01_Holster_Cloze_Elide.pdf

Meyer, J. P. (2014). *Applied Measurement with JMetrik*. Routledge. https://books.google.co.jp/books?id=Xd-8mQEACAAJ