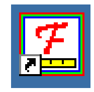
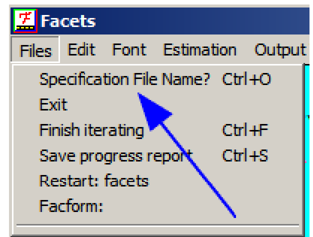
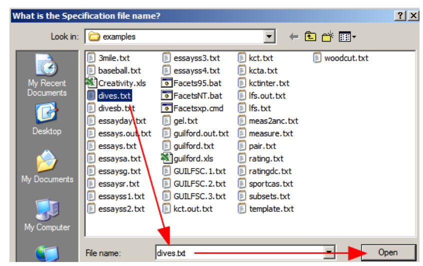
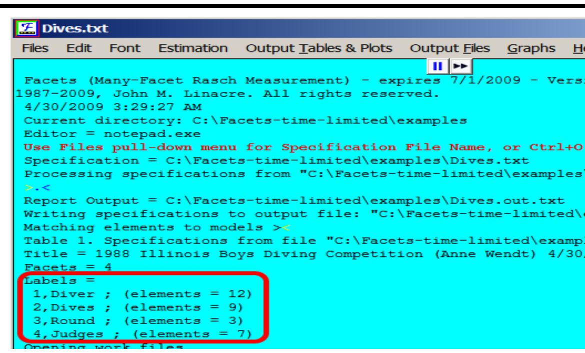
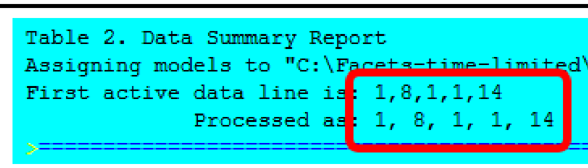
| # | Many-Facet Rasch Measurement : Facets Tutorial<br>Mike Linacre - 1/2012 | |
|---|---|---|
| 1. | **Tutorial 3. Estimation and interactions**<br>***The adventure continues!***<br>• Estimation methods<br>• How iterative estimation works<br>• Interactions and differential item functioning (DIF)<br>This tutorial builds on Tutorials 1 and 2, so please go back and review them when you need to. | |
| 2. | **A. Facets Analysis Window: Illinois High School Diving Competition** | |
| 3. | Let's launch Facets again |  |
| 4. | We'll start this Tutorial by looking at some real-life, messy data<br>Click on "Files"<br>Click on "Specification File Name?" |  |
| 5. | Click on "Dives.txt" and "Open"<br>or Double-Click on "Dives.txt"<br>"Extra Specifications" - click on "OK"<br>"What is the Report Output file name" - click on "Open" |  |
| 6. | Click on the *Facets analysis* window, and scroll back up to the top.<br><br>The analysis starts by processing the specifications, and then briefly reports them at the same time as the Output Tables are being written.<br><br>We see this is a 4-facet analysis.<br>The facets are: 1-Diver, 2-Dives, 3-Round, 4-Judges |  |
| 7. | The *Facets* reads in the data. The first data line is shown on the screen, together with how it appears to *Facets*. If these lines look incorrect in your analysis, then inspect your data file. |  |

**8.** Scroll down to
Table 2: Look closely, Facets reports 38 lines in the data file and 238 ratings.

There are two model statements. They are matched with the data starting with the model at the top of the list each time:

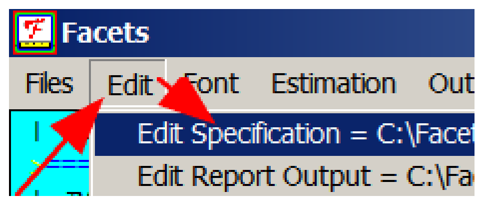
**Model = 7,8,?,?,M** means "diver 7 on dive 8 in any round with any judge is rated on rating scale M". "M" means "treat as missing data". So the 7 data points matching this model will be treated as missing data.

**Model = ?,?,X,?,DOUBLEPOINTS,1** means "any diver with any dive, ignoring the round, with any judge are rated on rating scale called "DOUBLEPOINTS" and each observation has a weight of 1. The remaining 231 ratings match this model.

```
Total lines in data file = 38
Total data lines = 38
Responses matched to model: 7,8,?,?,M = 7  ; treated as missing
Responses matched to model: ?,?,X,?,DOUBLEPOINTS,1 = 231
    Total non-blank responses found = 238
         Number of blank data lines = 1
Valid responses used for estimation = 231
```

If this was the order:
**Model =**
**?,?,X,?,DOUBLEPOINTS,1**
**7,8,?,?,M**
*****
Then all the data would match the "DOUBLEPOINTS" statement, and none of the data would be treated as missing.

**9.** Click on the Facets menu bar.
Click on "Edit Specification file"




**10.** In the NotePad window, scroll down to **Models=**
You can see the model specifications corresponding to those in the *Facets analysis* window.
"M" means "treat as missing data"
"DoublePoints" is what we have called our rating scale.

```
Models =          ; these
 7,8,?,?,M        ; make
 ?,?,X,?,DoublePoints
*
```

**11.** Scroll down to the element list after Labels=

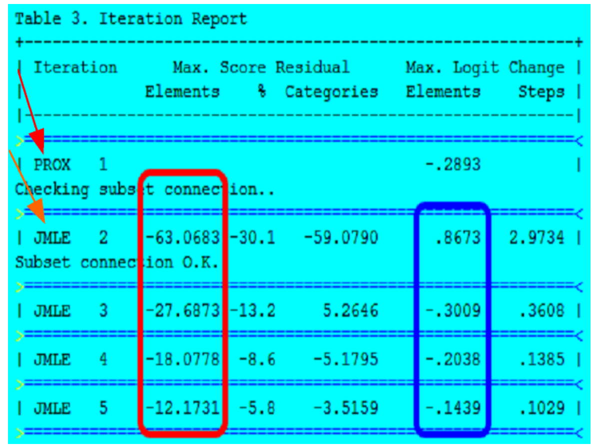
1, Diver
1, Marty Turek 292.85 , 2.08
1 is the element number
Marty Turek 292.85 is the element label. 292.85 is here for reference. It was Marty's score in the preliminary competition.
2.08 is a logit value which is used as the initial dive measure estimate. This value is optional and was used to speed up estimation when computers were slow.

For the dives,
1, 1.4 , -0.60
1.4 is the element label for element 1. It is the official "weight" of the dive in the raw-score scoring procedure used at the diving competition.

```
Labels=
1,Diver
;   diver - previous scores
 1, Marty Turek       292.85  ,  2.08    ; logit starting values
 2, Tom Wright        279.95  ,  1.63
 3, Mike Gotkowski    249.9   , -0.33
 4, Matt Paulson      244.55  ,  1.10
 5, Scott Ternovits   252.8   , -1.28
 6, Ross Moyer        243.4   , -0.15
 7, Curt Billings     266.25  ,  1.54
 8, Steve Hutchings   267.15  ,  0.82
 9, Larry Kirk        258.35  , -0.38
10, Kurt Becker       284.4   , -0.36
11, Lance Kleffman    259.6   , -0.81
12, Bryan Hanania     251.15  , -1.19
*
2,Dives
 1, 1.4                       , -0.60    ;starting values
 2, 1.7                       , -0.98
 3, 1.8                       ,  0.35
 4, 2.2                       , -0.16
 5, 2.3                       , -0.51
 6, 2.4                       ,  0.44
 7, 2.5                       ,  0.46
 8, 2.6                       ,  0.99
 9, 2.4?                                 ; unclear what dive this was
*
```

| | | |
|---|---|---|
| **12.** | Scroll down to **Data=**<br>1,8,1,1,14 is diver 1 (Marty Turek) doing dive 8 (labeled 2.6) in round 1 and rated by judge 1.<br>Judge 1 awarded a “7”, but, because there were half-score-points, such as 5.5, this is entered as 14. *Facets analyses integer data, so all the observed ratings were multiplied by 2.*<br>Most data lines show judges 1-7, but some don’t. This is because the data were recorded by one of the diving coaches, who was busy coaching his own divers at the same time. So some observations are missing. This is fatal for some types of analysis, but not for Rasch. The Rasch measures are based on the active data. There is no need to fill in (impute) missing data, or to do any item or person deletion in order to make the data “complete”. | `Data=`<br>`1,8,1,1,14`<br>`1,8,1,2-4,11,12,12`<br>`1,1,2,1-7,16,13,13,16,14,13,14`<br>`2,8,2,1-7,12,12,10,11,10,11,13`<br>`3,1,1,1-4,10,11,10,11`<br>`3,6,2,1-7,11,10,09,08,09,09,11`<br>`4,3,1,1-7,12,13,13,13,10,12,12`<br>5,6,7 are *Missing data* |
| **13.** | Further down the data file, all seven judges awarded 0 to a dive. This was because the diver performed the wrong dive. So, when measuring the divers we need to include these data, but for other purposes, such as evaluating judge behavior, these observations should be made missing. Hence the “M” model we saw above. The “M” model matches these, and only these, observations. *Do you see how this works??* | `. . .  . . . . . . .`<br>`; these next observations are made missing - wrong dive!`<br>`7,8,3,1-7,00,00,00,00,00,00,00`<br><br>These observations produce big misfit in the diver, dives, rounds and judges. Only use these ratings for reporting the final diver measures. |
| **14.** | Scroll to the **Models=** and **Rating Scale=** specifications:<br>The rating scale is called “DoublePoints” - as usual, **DoublePoints** is defined to be a rating with:<br>**R20** - a rating scale with highest category 20. | `Rating scale = DoublePoints,R20,Keep`<br>`0  = 0.0`<br>`10 = 5.0       ; 5 on the original s`<br>`20 = 10.0`<br>`*` |
| **15.** | **Keep** - keep in the rating scale any unobserved intermediate categories. These are called *incidental or sampling zeroes.*<br>The default option is to squeeze out unobserved intermediate categories *(structural zeroes),* and then renumber the remaining categories sequentially. For example: tennis scores are 0-15-30-40-(50 advantage)-(60 game). Facets automatically converts these to 0, 1, 2, 3, 4, 5 points for analysis. | |
| **16.** | 0=0.0<br>10=5.0<br>20=10.0<br>* | In the DoublePoints rating scale, there are 21 categories, 0-20, and we’ve chosen to label some of them. Category 10 is labeled “5.0”. This is because the original ratings of the dives were numbers like 3.5 and 5.5 out of 10. Facets expects integer ratings. So all the ratings have been multiplied by 2, but we want to remember what the original ratings were, so we use the “Rating Scale=” category labels to do that. |

**17.** Back in the *Facets analysis* window, we have reached **Table 3**, the estimation Iteration Report. Facets goes through the data numerous times in order to estimate the abilities, difficulties, severities, etc. Each traversal of the data is called an iteration. It is shown by >=====<
red arrow: PROX is the "normal approximation algorithm". This speedy but somewhat inexact algorithm is used to obtain initial measures. PROX approximates the observed distributions by normal distributions These measures become the starting values for values for the more exact, but slower,
orange arrow: JMLE algorithm.



```
Table 3. Iteration Report
+---------------------------------------------------------------+
| Iteration        Max. Score Residual       Max. Logit Change  |
|                 Elements   %  Categories  Elements     Steps  |
|---------------------------------------------------------------|
>=============================================================<
| PROX    1                                    -.2893             |
Checking subset connection..
>=============================================================<
| JMLE    2    -63.0683 -30.1   -59.0790        .8673    2.9734 |
Subset connection O.K.
>=============================================================<
| JMLE    3    -27.6873 -13.2     5.2646       -.3009     .3608 |
>=============================================================<
| JMLE    4    -18.0778  -8.6    -5.1795       -.2038     .1385 |
>=============================================================<
| JMLE    5    -12.1731  -5.8    -3.5159       -.1439     .1029 |
>=============================================================<
```

**18.** JMLE, "joint maximum likelihood estimation", also called UCON, "unconditional maximum likelihood estimation", is robust against missing data and non-normal Rasch-measure distributions. Its essential idea is that for each Rasch measure (element measure, or rating scale Rasch-Andrich threshold), the estimated measure is the value for which **the observed raw score the element's observations is the same as the expected raw score based on the element's estimated measure.**

The JMLE estimate is also the estimate for which the likelihood of the observed data occurring is at its maximum. So that we can say:
Your raw score → Your estimated measure → Your most likely raw score is your observed raw score

When observed raw score for person n = the expected raw score for person n, then we have estimated person n's ability:

$$\left( \sum_n X_{ni} = \sum_n E_{ni} \right) \Rightarrow \hat{B}_n$$

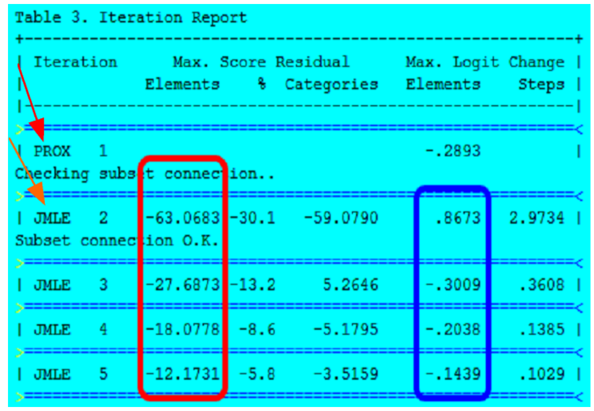
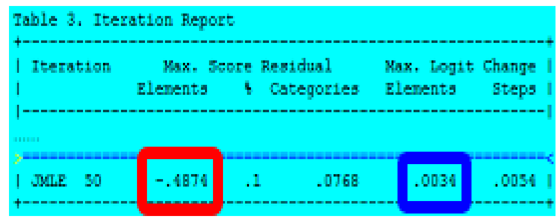
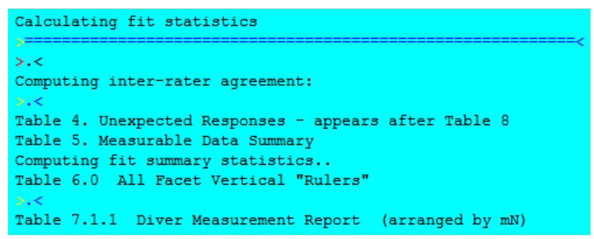
We can update the estimates using the Newton-Raphson estimation equation:

$$\hat{B}_n' = \hat{B}_n + \left( \sum_n X_{ni} - \sum_n E_{ni} \right) \Big/ \sum_n V_{ni}$$

Revised estimate = Previous estimate + (observed score - expected score)/(variance)

**19.** *Here is how JMLE works .... Fortunately we don't do this. The computer does it for us!*

1. Compute the raw scores (marginal scores) for all the elements (persons, items, etc.). Let's say that the data are dichotomous (0-1), and element 10 has a score of 15 for 20 observations.
2. All the element measures (abilities, difficulties, etc.) are assigned estimated starting values. Let's say 0 logits.
3. The expected value of all observations is computed based on the current element measures. Let's say the data are dichotomous (0-1). Then all the expected values are 0.5. The model variance of each observation is also computed, it is 0.5*0.5 = 0.25 per observation.
4. The expected raw scores and summed variances for all the elements are computed. Then element 10's 20 observations have an expected score of 0.5 * 20 = 10, and a summed variance of 0.25 * 20 = 5
5. Compute a better estimate, B', of the current measure B:
   B' = B + (observed raw score - expected raw score)/(summed variance)
   For element 10,
   B' = 0 + (15 - 10)/5 = 1
   Do this for all the other elements.
6. We now have a better set of estimates. Return to 2 and redo the computation, until the change in estimates is very small.

<table>
<tr>
<td>20.</td>
<td>After each iteration through the data, Winsteps reports<br>Red box: the biggest (furthest from zero) difference between the observed raw score for any person or item and the corresponding expected raw score. We expect this value to reduce to less than .5 score points, the smallest difference visible in the data.<br>Blue box: the biggest (furthest from zero) difference between the current estimate and the previous estimate for any person or item. We expect this value to reduce to less than .01 logits, so that changes are too small to change the output Tables.</td>
<td>
</td>
</tr>
<tr>
<td>21.</td>
<td>Iteration 50 is the last iteration, the last traversal of the data before convergence. We can see that the values in its last line are small.<br>Red box: the biggest difference between the observed and expected scores for any element is -.4874, which is closer to zero than 0.5<br>Blue box: The biggest change in any estimate is .0034, too small to change the measures that we see in the output Tables.</td>
<td>
</td>
</tr>
<tr>
<td>22.</td>
<td>After the measures are estimated, the fit statistics are computed, and also the rater agreement summaries.<br><br>The Tables are now written to the Report Output file.</td>
<td>
</td>
</tr>
<tr>
<td>23.</td>
<td></td>
<td></td>
</tr>
</table>

**24.** **B. Output Tables: Illinois High School Diving Competition**

**25.** Here is **Table 6** in the Report Output file, dives.out.txt. (You know how to get there ...).
*NotePad: reduce the size of Table 6 with: "Format" "Font" "Size" 8 OK*

Do you notice the + and - at the top of the columns?
+ means positive facet: more score = more measure
- means negative facet: more score = less measure
Use Positive= or Negative= to set these directions.

Red box: We can give the diving coaches some useful information from this analysis. The dives are labeled by their official difficulty weights, such as 2.6. We can see it is the label of the most difficult dive, highest in the "-Dives" column. The easiest dive, labeled 1.7, is near the bottom. These weights are used in the diving competition to multiply the diver's scores. They are intended to adjust for dive difficulty

*Here is our advice to the coaches:*
**Green circle:** the 2.3 dive has a big weight, but is relatively easy. Encourage your diver to perform this dive!
**Orange circle:** The 1.8 dive has a low weight, but is relatively difficult. Avoid it!
*???Why doesn't every diving coach do this analysis???*

```
+--------------------------------------------------------------------------+
|Measr|+Diver                                        |-Dives   |-Judges|DOUBL|
|-----+----------------------------------------------+---------+-------+-----|
|  2  +                                              +         +       +(16) |
|     |                                              |         |       |     |
|     | Marty Turek     292.85                       |         |       | --- |
|     |                                              |         |       |     |
|     |                                              |         |       |     |
|     |                                              |         |       |     |
|     |                                              |         |       |     |
|     | Tom Wright      279.95                       |         |       | 13  |
|     | Curt Billings   266.25                       |         |       |     |
|     |                                              |         |       |     |
|  1  +                                              + 2.6     +       +     |
|     |                                              |         |       |     |
|     | Matt Paulson    244.55                       |         |       | --- |
|     |                                              |         |       |     |
|     |                                              |         |       |     |
|     | Steve Hutchings 267.15                       | 2.5     |       |     |
|     |                                              | 1.8 2.4 |       | 12  |
|     |                                              |         | 5     |     |
|     |                                              |         |       |     |
|     |                                              |         | 3  6  | --- |
*  0  *                                              *         * 2  4  *     *
|     |                                              |         |       |     |
|     |                                              | 2.2     | 1     | 11  |
|     |                                              |         |       |     |
|     |                                              |         | 7     |     |
|     | Ross Moyer      243.4                        | 2.3     |       | --- |
|     |                                              | 1.4     |       |     |
|     | Kurt Becker     284.4   Larry Kirk    258.35  Mike Gotkowski  249.9 |       |     |
|     |                                              |         |       | 10  |
|     |                                              |         |       |     |
| -1  +                                              + 1.7     +       +     |
|     |                                              |         |       | --- |
|     | Lance Kleffman  259.6                        |         |       |     |
|     |                                              |         |       |  9  |
|     |                                              |         |       |     |
|     | Bryan Hanania   251.15                       |         |       |     |
|     |                                              |         |       | --- |
|     | Scott Ternovits 252.8                        |         |       |     |
|     |                                              |         |       |     |
|     |                                              |         |       |  8  |
| -2  +                                              +         +       + (6) |
|-----+----------------------------------------------+---------+-------+-----|
|Measr|+Diver                                        |-Dives   |-Judges|DOUBL|
+--------------------------------------------------------------------------+
```

**26.** **Table 8:** How did the long rating-scale function? Scroll down to Table 8. You will see the rating scale functioned amazingly well.

**Green box:** The Average Measures advance much in line with Rasch expectations (blue box).

Red box: The only noticeable misfit (OUTFIT MnSq = 2.0) is for category 15 with only 1 observation.

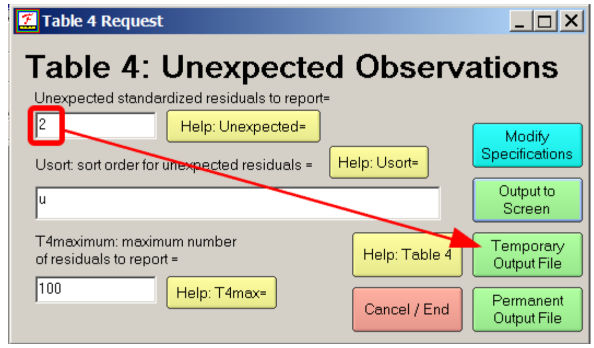
*Notice the range of Category Scores: 6 - 16.* Categories below 6 and above 16 were not observed, so they cannot be estimated.

If we must include unobserved categories in our analysis, then there are techniques to do this using dummy (artificial) data or imputed rating-scale structures, such as binomial trials.

| DATA Category Score | Counts Used | % | Cum. % | QUALITY CONTROL Avge Meas | Exp. Meas | OUTFIT MnSq | RASCH-A Thresh Measure |
|---|---|---|---|---|---|---|---|
| 6 | 10 | 4% | 4% | -2.00 | -1.99 | 1.0 | |
| 7 | 13 | 6% | 10% | -1.82 | -1.86 | .8 | -2.18 |
| 8 | 23 | 10% | 20% | -1.67 | -1.62 | 1.1 | -2.32 |
| 9 | 25 | 11% | 31% | -1.10 | -1.16 | 1.2 | -1.49 |
| 10 | 33 | 14% | 45% | -.70 | -.57 | .7 | -1.14 |
| 11 | 34 | 15% | 60% | -.10 | -.12 | 1.1 | -.36 |
| 12 | 48 | 21% | 81% | .31 | .21 | .7 | -.30 |
| 13 | 29 | 13% | 93% | .52 | .55 | 1.1 | .87 |
| 14 | 13 | 6% | 99% | 1.00 | 1.06 | 1.0 | 1.58 |
| 15 | 1 | 0% | 99% | 1.02 | 1.80 | 2.0 | 3.98 |
| 16 | 2 | 1% | 100% | 2.49 | 2.30 | .6 | 1.36 |

**27.** Let us investigate the observation in category 15.
**Table 4 (after Table 8):** in the Output Tables says: “No unexpected observation with StRes >= 3”, so we will command *Facets* to produce Table 4 with **StRes >= 2.**

***Facets* Analysis window** menu bar:
Click on **Output Tables & Plots**
Click on **Table 4: Unexpected Observations**
Unexpected standardized residuals to report = **2**
Click on **“Temporary Output File”**

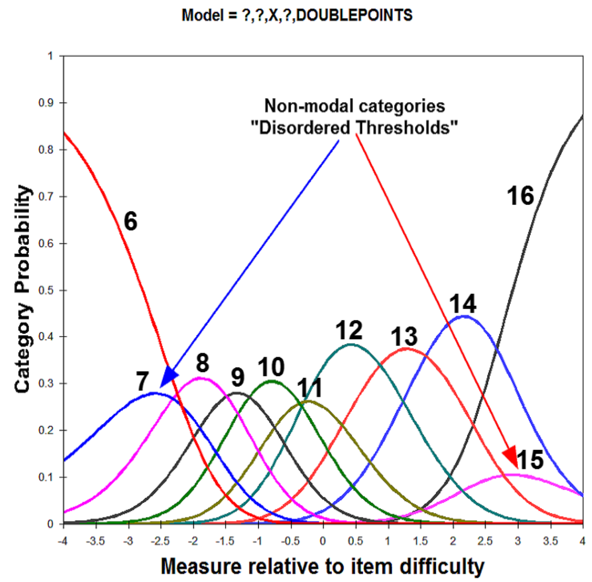
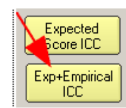
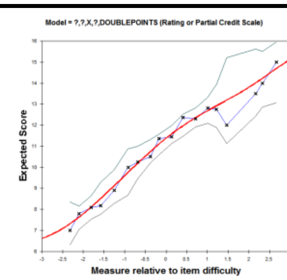
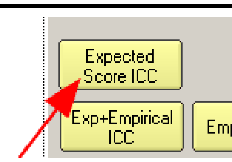



**28.** **Table 4** displays in a NotePad window.

Red box: there is the observation in category 15! It was expected to be 12.7 (which would be observed as 13). 15 was a surprisingly **high** rating for Steve Hutchings. Its residual = (observed - expected) = 15 - 2.7 = 2.3 is in the Resd column. Its unexpectedness is in the StRes column. This is the “standardized residual” of 2.2, which is as unlikely as a value of 2.2 is on a unit-normal distribution (Tutorial 2 Appendix 1).

```
1988 Illinois Boys Diving Competition (Anne Wendt) 4/30/2009 3:29:27 AM
Table 4.1 Unexpected Responses (7 residuals sorted by u).

+---------------------------------------------------------------------+
| Cat  Score  Exp.  Resd StRes| Nu Diver                     N Div N J |
|-----------------------------+----------------------------------------|
|  8      8   11.4  -3.4 -2.8 |  9 Larry Kirk       258.35 5 2.3 1 1 |
|  9      9   12.0  -3.0 -2.6 |  8 Steve Hutchings 267.15 6 2.4 7 7 |
|  8      8   11.0  -3.0 -2.3 |  9 Larry Kirk       258.35 5 2.3 4 4 |
| 11     11    8.1   2.9  2.2 |  3 Mike Gotkowski   249.9  8 2.6 3 3 |
|  9      9   11.7  -2.7 -2.2 |  4 Matt Paulson     244.55 7 2.5 3 3 |
| 15     15   12.7   2.3  2.2 |  8 Steve Hutchings 267.15 4 2.2 7 7 |
|  6      6    9.0  -3.0 -2.2 | 10 Kurt Becker      284.4  8 2.6 7 7 |
|-----------------------------+----------------------------------------|
| Cat  Score  Exp.  Resd StRes| Nu Diver                     N Div N J |
+---------------------------------------------------------------------+
```

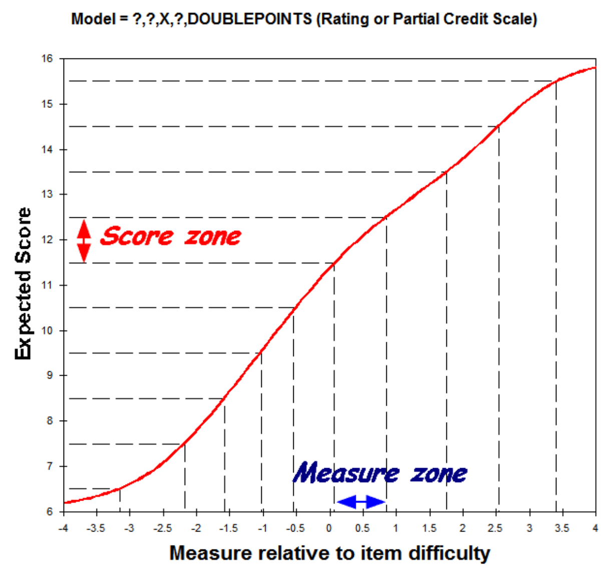
| 29. | **C. Graphs: Illinois High School Diving Competition** | |
|---|---|---|
| **30.** | Let's look at a picture of what Table 8 means ....<br>On the Facets menu bar, click on "**Graphs**".<br><br>The Rasch-model category probability curves look the desired "range of hills", with only a slight problem due to low category-probabilities ("Rasch-Andrich threshold disordering") at each end (blue and red arrows).<br><br>Overall, this is a remarkably good picture of probability curves, especially considering that it came from a long rating scale with thin data.<br><br>The low-probability categories are probably an accident of this sample, never to be repeated in exactly this way. |  |
| **31.** | Take a glance at the expected and empirical Item Characteristic Curves (ICCs) |  |
| **32.** | The empirical curve (thin blue line with x's) tracks the model, continuous red line, closely for most of the operational range of the scale. It is well within the confidence bands (light lines).<br><br>Inferences from rating-scale categories to measure are well-supported by these data. |  |
| **33.** | Click on the "Expected Score ICC" |  |

**34.** This shows the "score-to-measure" ogive for an individual item.

Red arrows: The score zone is between expected half-rating-points. In this example, 11.5 to 12.5. We can imagine people's performances advancing smoothly up the rating scale. But fractional score points can't be observed, so this is the score zone corresponding to a rating of "12".

Blue arrows: The matching measure zone goes from .08 to .84 logits. We will see these exact numbers in Table 8 - the next picture.

*So here is the logic:* for measures in the zone .08 to .84 relative to item difficulty, we expect the average rating of our sample to be in the range 11.5 to 12.5. But fractional ratings can't be observed, so we expect the observed ratings to be 12 or near to 12.




**35.** Here is Table 8 again, if you compare the dashed lines in the plot above with this Table, you can see where these numbers come from.

Red box: the category numbers

Blue box: measures at -0.5 score-points

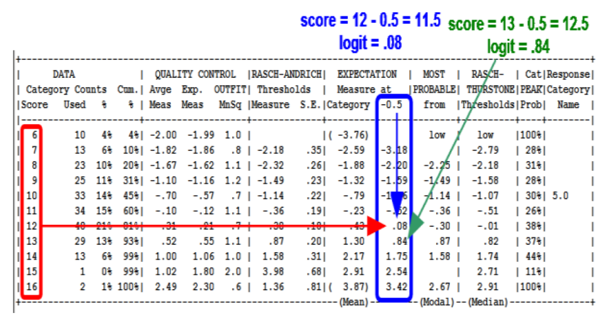
Blue arrow for category 12 - 0.5 = 11.5 = .08 logits

**Green arrow** for category 13 - 0.5 = 12.5 = .84 logits

Score zone = 11.5 - 12.5 score-points

Measure zone = .08 - .84 logits

*Can you match these numbers to the ogive in #34?*



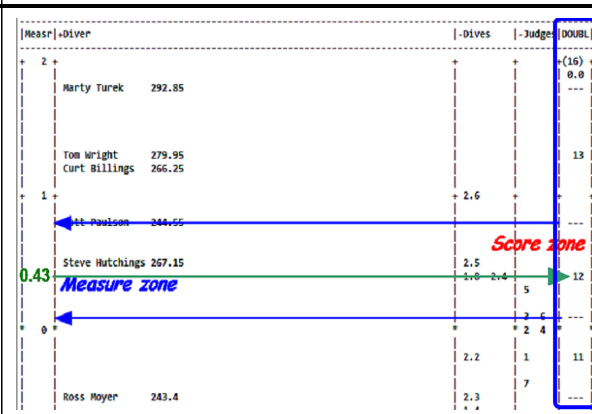

| DATA Category Score | Counts Used | % | Cum. % | QUALITY CONTROL Avge Meas | Exp. Meas | OUTFIT MnSq | RASCH-ANDRICH Thresholds Measure | S.E. | EXPECTATION Measure at Category | -0.5 | MOST PROBABLE from | RASCH-THURSTONE Thresholds | Cat PEAK Prob | Response Category Name |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 6 | 10 | 4% | 4% | -2.00 | -1.99 | 1.0 | | | ( -3.76) | | low | low | 100% | |
| 7 | 13 | 6% | 10% | -1.82 | -1.86 | .8 | -2.18 | .35 | -2.59 | -3.18 | | -2.79 | 28% | |
| 8 | 23 | 10% | 20% | -1.67 | -1.62 | 1.1 | -2.32 | .26 | -1.88 | -2.20 | -2.25 | -2.18 | 31% | |
| 9 | 25 | 11% | 31% | -1.10 | -1.16 | 1.2 | -1.49 | .23 | -1.32 | -1.59 | -1.49 | -1.58 | 28% | |
| 10 | 33 | 14% | 45% | -.70 | -.57 | .7 | -1.14 | .22 | -.79 | [illegible] | -1.14 | -1.07 | 30% | 5.0 |
| 11 | 34 | 15% | 60% | -.10 | -.12 | 1.1 | -.36 | .19 | -.23 | -.52 | -.36 | -.51 | 26% | |
| 12 | 40 | 21% | 81% | .31 | .21 | .7 | .30 | .10 | .43 | .08 | -.30 | -.01 | 38% | |
| 13 | 29 | 13% | 93% | .52 | .55 | 1.1 | .87 | .20 | 1.30 | .84 | .87 | .82 | 37% | |
| 14 | 13 | 6% | 99% | 1.00 | 1.06 | 1.0 | 1.58 | .31 | 2.17 | 1.75 | 1.58 | 1.74 | 44% | |
| 15 | 1 | 0% | 99% | 1.02 | 1.80 | 2.0 | 3.98 | .68 | 2.91 | 2.54 | | 2.71 | 11% | |
| 16 | 2 | 1% | 100% | 2.49 | 2.30 | .6 | 1.36 | .81 | ( 3.87) | 3.42 | 2.67 | 2.91 | 100% | |
| | | | | | | | | | (Mean) | | (Modal) | (Median) | | |

**36.** And here **Table 6**, which we saw in #24,

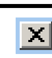
Blue box: it shows the same score-zones ....

Blue arrows: The measure-zone on the logit scale matches with the score-zone on the rating scale. The "---" in the rating-scale blue box indicates a .5 score point.

Green arrow: The integer ratings, e.g., "12", are shown where 12 is the expected score. This is 0.43 logits, which is the "EXPECTATION Measure at Category" in Table 8 in #35




**37.** Close all windows ...

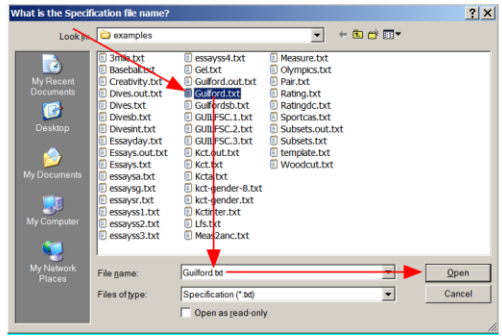


**38.**

**39.** **D. Bias/Interaction Analysis: Guilford's Creativity Ratings**

**40.** In *Tutorial 2,* we discovered that J.P. Guilford's data in Guilford.txt has serious flaws. Let's investigate those further. This is where *Facets* can really help us ...

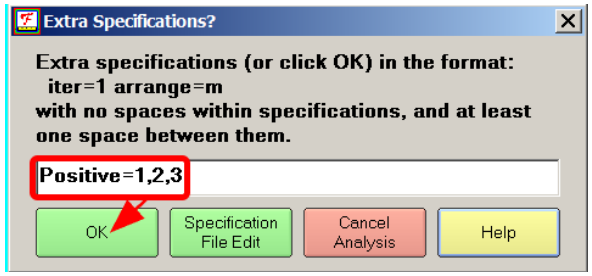
Launch *Facets*
On the *Facets* menu bar,
Click on "Specification File Name?"
Double-click on "**Guilford.txt**"




**41.** We will make this section easier to understand if all three facets are positive (ability, leniency, severity):

"Extra Specifications?"
Positive=1,2,3
Click "OK"

"What is the Report Output Filename?" Click "Open"
Estimation is performed and the Report Output file is written and displayed.




**42.**

**43.**

Look at **Table 1 in the Report Output File**
There are 3 facets: Senior Scientists ( facet 1, judges), Junior Scientists (facet 2, examinees) and Traits (facet 3, items)
Red arrow: the facets are positively oriented.
Red box: The measurement model is:
Model = ?B, ?B, ?, CREATIVITY, 1
We know that ?,?,? means "any element of facet 1 can combine with any element of facet 2 and any element of facet 3."

```
; Data specification
Facets = 3
Non-centered =  1
Positive =  1, 2, 3
Labels =
 1,Senior scientists ; (elements = 3)
 2,Junior Scientists ; (elements = 7)
 3,Traits ; (elements = 5)
Model = ?B,?B,?,CREATIVITY,1
Rating (or other) scale = CREATIVITY,R9,General,Ordinal
```

**44.** **Table 6** in the Report Output File (or the "Output Tables" menu)

Red box: Notice that the Senior Scientist (Facet 1, judge) with the highest measure is Cavendish.
Red arrow: positive facet. High measure = hig score.
Cavendish is the most lenient. Brahe is the most severe.

```
Vertical = (2N,3A,2*,1A,1A,S) Yardstick (columns lines low high extreme)= 0,10,-1,1,End
+-------------------------------------------------------------------------------------
|Measr|+Junior Scientists|+Traits                 |+Junior Scientists|+Senior scientists
|-----+------------------+------------------------+------------------+-----------------
|   1 +                  +                        +                  +
|     |                  |                        |                  |
|     |                  |                        |                  |
|     |                  |                        |                  |
|     | 2                |                        | *                |
|     |                  |                        |                  |
|     | 5                |                        | *                |
|     | 7                | Attack      Daring     | *                |
|     |                  |                        |                  |
|     |                  | Basis                  |                  | Cavendish
*   0 *                  *                        *                  * Avogadro
|     | 1                |                        | *                |
|     | 3                | Clarity                | *                | Brahe
|     |                  |                        |                  |
|     |                  |                        |                  |
|     | 4                | Enthusiasm             | *                |
|     | 6                |                        | *                |
|     |                  |                        |                  |
|     |                  |                        |                  |
|     |                  |                        |                  |
|  -1 +                  +                        +                  +
|-----+------------------+------------------------+------------------+-----------------
|Measr|+Junior Scientists|+Traits                 | * = 1            |+Senior scientists
+-------------------------------------------------------------------------------------
```

**45.** Model = ?B, ?B, ?, CREATIVITY, 1

**“?B,?B”** means that, after performing the main estimation and reporting it, perform a **Bias/Interaction analysis** between facet 1 and facet 2.
When the bias is between items and groups of persons, this is equivalent to investigating *Differential Item Functioning, DIF.*

**46.** The analytical model is:
1. Model = ?, ?, ?, CREATIVITY
Estimate the measures, the main effects.

2. Compute the residuals = observed - expected

3. B, B: Estimate the bias interactions, the secondary effects, between “Senior Scientists” (judges, $C_j$) and “Junior Scientists” (examinees, $B_n$), $C_{jn}$, based on the the residuals, $R_{nij}$, in the cluster of responses where judge j rated examinee n.

1. $\{X_{nij}\} \Rightarrow \hat{B}_n + \hat{D}_i + \hat{C}_j - \hat{F}_k$

2. $\{X_{nij} - E_{nij}\} \Rightarrow R_{nij}$

3. $\{R_{nij}, \hat{B}_n, \hat{D}_i, \hat{C}_j, \hat{F}\} \Rightarrow \hat{C}_{jn}$

**47.** Scroll down to **Table 9**. This reports the estimation of the bias/interaction. You can usually **ignore** this Table and also **Table 11**. They are reported so that you can verify that the estimation process has progressed correctly, if you ever need to.

**Table 10** is not usually meaningful, so it is not reported.

```
Ratings of Scientists (Psychometric Methods p.282 Guilford 19
Table 9.1  Bias Iteration Report.

Bias/Interaction: 1. Senior scientists, 2. Junior Scientists

There are empirically 21 Bias terms
+--------------------------------------------------------+
| Iteration      Max. Score Residual      Max. Logit Change |
|              Elements     %  Categories  Elements    Steps |
|--------------------------------------------------------|
| BIAS    1     -12.9867 -144.                -1.0000       |
| BIAS    2      -2.2472 -25.0                 -.2496       |
| BIAS    3       -.1364  -1.5                 -.0172       |
| BIAS    4       -.0006    .0                 -.0001       |
+--------------------------------------------------------+
```

**48.** **Table 12** reports the bias/interaction terms graphically. This display can be useful if you have many interaction terms to report.

```
Ratings of Scientists (Psychometric Methods p.282 Guilford 1954)  04-19-200
Table 12.1  Bias/Interaction Summary Report.

Bias/Interaction analysis specified by: 1. Senior scientists, 2. Junior Sci

Bias/Interaction Size:
               1            1  2   311 12  1 2 12     3
 +-------------------+Q---------S--------M--------S---------Q+
-2                  -1                   0                   1

Bias/Interaction Significance:
      1                  1 11     211 1  12   1   2   2 1    11   1
 +---------+-Q-------+-----S---+---------M---------+---S-----+-------Q-+
-4         -3        -2        -1        0         1         2         3
```

| | |
|---|---|
| **49.** | Bias/Interaction Size: |

```
                                1
                 12213457677888808866553422222
  2  21 1122472485329801372717797531096665273067453 2221111 1   1
 +---------+----Q----+-S-------M------S--+----Q----+---------+---------+
-3        -2        -1         0         1         2         3      4

Bias/Interaction Significance:
                             11111
                   1213358889100108775433222
     11 1 1 112764447736758335576842343950948933312221111  1     1
 +-----+-----+----Q+-----S-----M-----S-----+Q----+-----+-----+-----+
-5    -4    -3    -2    -1     0     1     2     3     4     5     6
```

**“Size” means “how big?” (in logits)**
**“Significance” means “how surprising?” (a probability shown as a unit-normal deviate)**
This summarizes the statistics shown in Table 13. The x-axis is the value. The vertical numbers are the count of interactions with that value. The numbers read downwards.
Red numbers in top vertical barchart: 107 interactions have a Bias/Interactions Size of 0.1 logits.
Red numbers in bottom vertical barchart: 115 interactions have a Bias/Interaction Significance of -0.3, which has a probability of about p=0.8 (double-sided). So these 115 interactions probably happened by chance.

**50.** **Table 13** displays the values of the bias/interaction terms. Table 13 contrasts local behavior (by a judge, examinee, item) with general behavior on the entire dataset. Only the most conspicuous interactions are displayed in Table 13.
Red box: Here we see that Judge **Brahe** and Examinee **David** have a noticeable interaction, as do Judge **Brahe** and Examinee **Edward**.

```
+------------------------------------------------------------------------------------------+
| Obsvd   Exp.  Obsvd  Obs-Exp| Bias  Model               |Infit Outfit|    Senior scientists  Junior Scientis |
| Score   Score Count  Average| Size   S.E.    t  d.f. Prob. | MnSq  MnSq | Sq N Senior sc  measr N Junior  measr |
|-----------------------------+--------------------------------+------------+---------------------------------------|
|   25    17.3    5     1.54|   .71   .29  2.42    4 .0726 |   .3    .3 | 11 2 Brahe      -.24 4 David     -.46 |
|   14    27.0    5    -2.60| -1.27   .36 -3.55    4 .0238 |   .7    .6 | 14 2 Brahe      -.24 5 Edward     .42 |
|-----------------------------+--------------------------------+------------+---------------------------------------|
```

**Arrange=** controls **t**he order of the elements.
**Zscore=** controls which elements are listed.

**51.** The first entry in Table 13 is for **Brahe and David.**
**Obsvd Count 5:** Brahe rated David 5 times (once on each of the 5 traits).
**Obsvd Score 25:** The observed sum of these 5 ratings is 25.
**Exp. Score 17.3:** But, based on Brahe’s overall severity David’s overall ability, we expected a summed score near to 17.3.
**Obs-Exp Average 1.54:** The observed ratings are ( 25 - 17.3 ) / 5 = 1.54 rating-points higher than we expected, on average.

```
+--------------------------------
| Obsvd    Exp.   Obsvd  Obs-Exp|
| Score    Score  Count  Average|
|-------------------------------+
|    25     17.3      5     1.54|
```

( Observed - Expected ) / Count = Average

**52.** Brahe has 1.54 rating-points local **leniency.**
**Bias size:** this leniency is .71 logits, with precision .29 logits.
**t:** So, a test of the hypothesis "this bias is due to measurement error", with a null hypothesis of "there is no statistically discernable bias in Brahe's ratings", has a *t* = 2.42
**d.f.:** the t-statistic has approximately "Obsvd Count-1", 5-1=4, d.f. so that
**Prob.:** p=.0726 (two-sided). The null hypothesis of "the same leniency" is not rejected at p<.05, but Brahe is locally more lenient than he usually is by .71 logits, which is almost statistically significant.

```
------------------------------------------
 Obs-Exp|  Bias  Model                   |
 Average|  Size   S.E.     t  d.f. Prob. |
--------+--------------------------------+
    1.54|    .71   .29   2.42    4 .0726 |
```

Either Brahe is locally more lenient by 0.71 logits, or David is locally more able by 0.71 logits, but here only the change in leniency makes sense.

**53.** In the second entry in Table 13, Brahe is 2.60 rating points less lenient (more severe) with Edward than he is overall. This is equivalent to 1.27 logits. This effect is statistically significant, p=.0238 (two-sided).

```
+-----------------------------------------------------------------------
| Obsvd   Exp.  Obsvd  Obs-Exp|  Bias  Model                          |
| Score  Score  Count  Average|  Size   S.E.      t   d.f.  Prob.     |
|-----------------------------+-----------------------------------------+
....|
|   14    27.0     5    -2.60|  -1.27   .36  -3.55     4  .0238     |
|-----------------------------+-----------------------------------------+
```

**54.** Combining these two interactions, Brahe is 1.98 logits more severe with Edward (-1.27) than with David (.71).

*Facets* does this computation for us in Table 14.

| Junior | Obs - Exp | Extra leniency | Brahe's overall leniency | Brahe's local leniency |
|---|---|---|---|---|
| David | 1.54 | .71 | -.24 | -.24 + .71 = 0.47 |
| Edward | -2.60 | -1.27 | -.24 | -.24 + -1.27 = -1.51 |
| Brahe's difference in leniency: | | | | 0.47 - -1.51 = 1.98 |

**55.** **Table 13** contrasts local behavior with general behavior.
**Table 14** contrasts local behavior of pairs of elements in the same facet.

**56.**

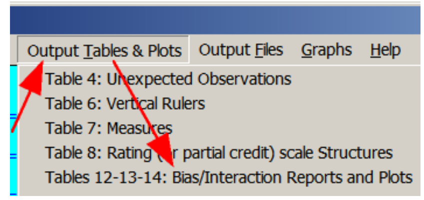
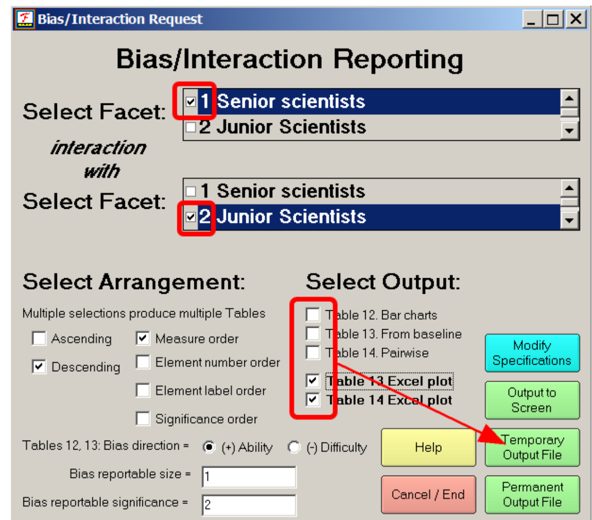
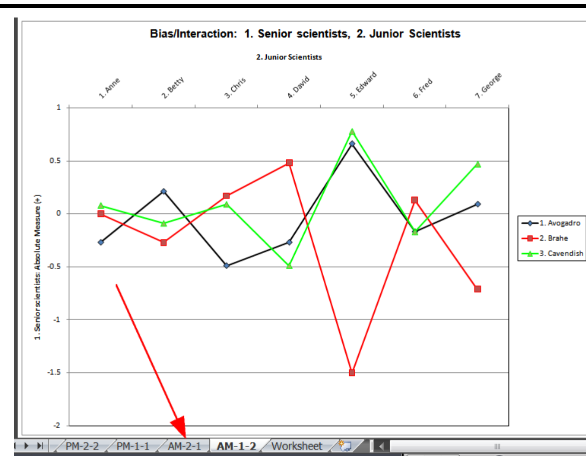
```
+--------------------------------------------------------------------------------------------------------------------+
| Target       | Target        Obs-Exp Context  | Target        Obs-Exp Context   | Target   Joint          Welch      |
| N Senior sc  | Measr  S.E.   Average N Junior | Measr  S.E.   Average N Junior  | Contrast  S.E.    t    d.f. Prob. |
|--------------+--------------------------------+---------------------------------+------------------------------------|
| 2 Brahe      |  -.48   .29      1.54 4 David  |  1.50   .36     -.66 5 Edward  |  -1.98    .46  -4.28    7 .0037 |
+--------------------------------------------------------------------------------------------------------------------+
```

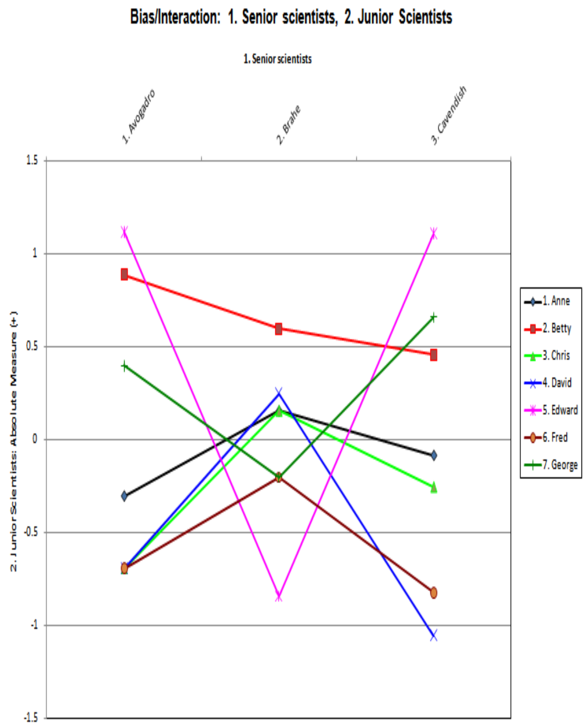
**57.** I've picked out the row in **Table 14** that corresponds to what we have just seen in **Table 13**:
Red boxes: The Target is Brahe. In the *Context* of David, Brahe (our *Target* judge) is .48 logits more lenient (giving higher ratings), but in the Context of Edward, Brahe is 1.50 logits less lenient (giving lower ratings).
Green boxes: Overall, Brahe is 1.98 logits more lenient with David than with Edward. This is the same as Table 13. This is a paired t-test, so we see that Brahe's change in leniency is highly significant, p =.0037.

**58.** Tables 13 and 14 can be difficult to understand because two things are usually going on at once, so, when interpreting the logit values, remember:

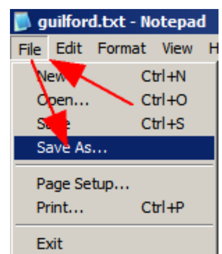
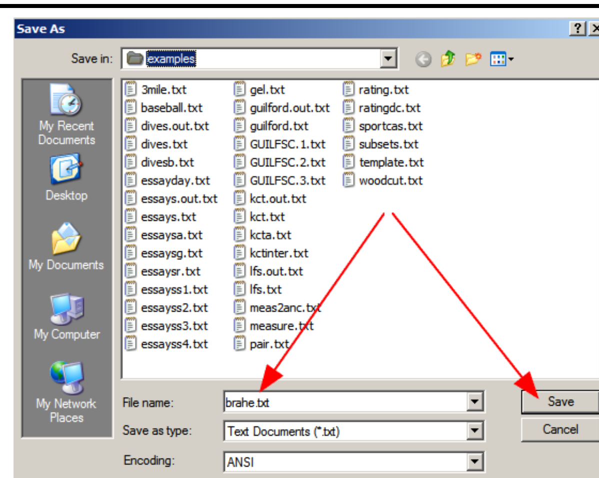
**Obs-Exp Average:**
**positive: higher ability, leniency, easiness**
**negative: higher severity, difficulty, lower ability**

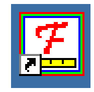
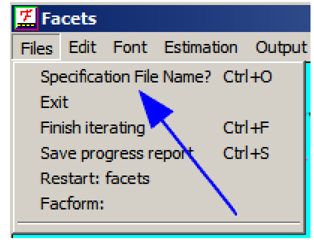
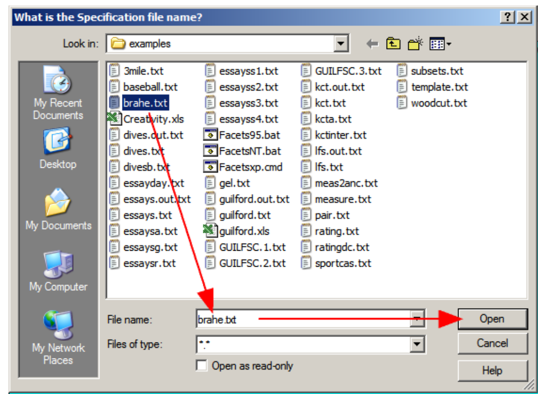
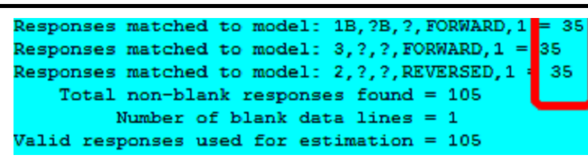
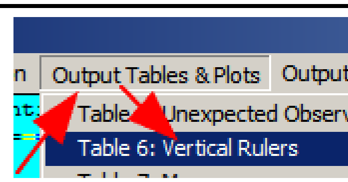
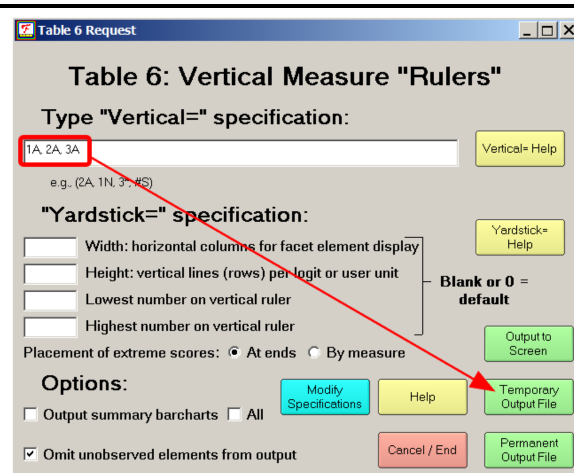
**59.**

| | | |
|---|---|---|
| **60.** | **E. Graphing Interactions with Excel** | |
| **61.** | Tables 13 and 14 contain a many details, and it is easy to become confused. So let's use Excel to plot the information for us. | |
| **62.** | *(If you need to restart Facets for Guilford.txt, remember* ***Positive=1,2,3*** *at Extra Specifications...)*<br>On the main *Facets* menu,<br>Click on "Output Tables & Plots".<br>Click on "Table 13-14: Bias/Interaction Report" |  |
| **63.** | In the "Bias/Interaction Request" dialog box,<br>Check the two facets whose interaction you want to investigate:<br>1. Senior scientists<br>2. Junior Scientists<br>Check "Table 13 Excel plot".<br>Check "Table 14 Excel plot".<br>Then click on "Temporary Output File"<br><br>Excel will also be launched ..... It may take a few seconds to display .. |  |
| **64.** | The Excel plot displays.<br><br>*"This is not the plot you are looking for. ... Move along now."*<br><br>Click on the bottom tab "AM-2-1"<br><br>If you cannot see "AM-2-1" then<br>Click on  to scroll to the right end of the tabs, then<br>Click on the bottom tab "AM-2-1" |  |

| | | |
|---|---|---|
| **65.** | *Plot AM-2-1:* "AM" means "Absolute Measure"<br>"2" means "the rows are facet 2 (junior scientists)<br>"1" means "the columns are facet 1 (senior scientists)<br><br>This is an amazing plot. It tells a remarkable story. Does it speak to you?<br><br>Columns: the three judges are shown from left-to-right.<br>Rows: the seven examinees are the seven colors.<br><br>The vertical axis is the examinees' absolute ability according to each of the three judges.<br>Mauve line: Look at Edward: according to Avogadro and Cavendish he is the most creative. According to Brahe he is the least creative!<br>Blue line: David has almost the opposite profile.<br><br>Can you see the pattern? Brahe's ordering of the examinees is almost the reverse of the ordering of Avogadro and Cavendish! |  |
| **66.** | This suggests that we could run two analyses:<br>1. Brahe's data by itself.<br>2. Then Avogadro's and Cavendish's data.<br>We can then compare the two sets of Junior Scientists' measures.<br>Red boxes: To omit the data for an element from an analysis, the fastest way is to comment out ";" the element in the *Labels=* specification:. | Labels=<br>1,Senior scientists<br>; 1=Avogadro ; Avogadro's data ignored<br>2=Brahe<br>; 3=Cavendish ; Cavendish's data ignored<br>*<br><br>Only Brahe's data will be analyzed .... |
| **67.** | | |

| 68. | **F. Specification File: Reversing Brahe's Ratings** | |
|---|---|---|
| 69. | *Let's try a different idea.* Suppose the problem is that Brahe misunderstood the rating scale and thought that 1 was the best, and 9 was the worst? Then, if we reverse Brahe's ratings, the whole analysis should make better sense! | |
| 70. | On the *Facets* menu bar,<br>Click on "Edit"<br>Click on "Edit Specification"<br>Guilford.txt displays in a NotePad edit window<br><br>We are going to edit the *Model* = and *Rating scale*= specifications | `Model = ?B,?B,?,Creativity ; judges, examinees :`<br>`; A bias/interaction analysis, ?B,?B,?, will`<br>`; senior scientists (judges,) and 2, junior :`<br>`; log(Pnijk/Pnijk-1) Bn - Di - Cj - Fk`<br>`; Bn = ability n, Di = difficulty i, Cj = Sev`<br>`; Pnijk = probability that child n on item i`<br>`Rating scale = Creativity,R9 ;Creativity is a rat:`<br>`1 = lowest ; name of lowest observed category`<br>`5 = middle ; no need to list unnamed categories`<br>`9 = highest ; name of highest observed category`<br>`*` |
| 71. | We want two models, one for Brahe and one for the other two judges.<br>And two rating scales, one for Brahe and one for the other two judges. | Model =<br>2, ?, ?, **Reversed** ; Judge 2, Brahe, uses the "Reversed" scale |
| 72. | Here is Brahe's rating scale. It uses the recoding option. See *Facets* Help for "Rating Scale=" for more information about this.<br>*Keep* is used because Brahe may not have used all the intermediate categories between 1 and 9, but we want to keep them all in the category ordering.<br><br>Red commas: , , , means that the values between the commas are to have their default values. In this case, "nothing". | Rating scale = **Reversed**, R9, Keep<br>1= nine , , , 9 ; 9 in data, recoded to 1.<br>2= eight , , , 8<br>3= seven , , , 7<br>4= six , , , 6<br>5= five , , , 5<br>6= four , , , 4<br>7= three , , , 3<br>8= two , , , 2<br>9= one , , , 1<br>* |
| 73. | For Avogadro and Cavendish, we could use the original *Models*= and *Rating scale*= , but I want to identify their rating.<br>I've added Bs in case we need to do a bias analysis. We only need to specify the Bs in one model statement. The Bs will apply to all the *Models*= | Model =<br>1B, ?B, ?, Forward ; Avogadro<br>3, ?, ?, Forward ; Cavendish |
| 74. | Add the Rating Scale= for Avogadro and Cavendish The name of the rating scale is "Forward." It is R9, a rating scale with highest category 9. It is "General" , so every reference to it in a *Models*= specification references the same rating scale structure. We want Cavendish and Avogadro to work together to define this rating scale structure. | Rating scale = Forward, R9, General, Keep<br>* |

| | | |
|---|---|---|
| **75.** | Edit guilford.txt<br><br>In place of<br><br>`Model = ?B,?B,?,Creativity ; judges, examinees and items produce ratings on "Creativity".`<br>`; A bias/interaction analysis, ?B,?B,?, will look for interactions between facets 1,`<br>`; senior scientists (judges,) and 2, junior scientists (examinees).`<br>`; log(Pnijk/Pnijk-1) Bn - Di - Cj - Fk`<br>`; Bn = ability n, Di = difficulty i, Cj = Severity j, Fk = Challenge k,`<br>`; Pnijk = probability that child n on item i is rated by judge j with score of k.`<br>`Rating scale = Creativity,R9 ;Creativity is a rating scale with possible categories 0 to 9`<br>`1 = lowest ; name of lowest observed category`<br>`5 = middle ; no need to list unnamed categories`<br>`9 = highest ; name of highest observed category`<br>`*` | Copy and paste into guilford.txt :<br><br>`Model =`<br>`1B, ?B, ?, Forward ; Avogadro`<br>`3, ?, ?, Forward ; Cavendish`<br>`2, ?, ?, Reversed ; Brahe`<br>`*`<br>`Rating scale = Forward, R9,General,Keep`<br>`*`<br>`Rating scale = Reversed, R9, Keep`<br>`1= nine , , , 9`<br>`2= eight , , , 8`<br>`3= seven , , , 7`<br>`4= six , , , 6`<br>`5= five , , , 5`<br>`6= four , , , 4`<br>`7= three , , , 3`<br>`8= two , , , 2`<br>`9= one , , , 1`<br>`*`<br><br>and also edit Positive= to become ...<br>`Positive = 1,2,3` |
| **76.** | In the *NotePad* Window:<br>Click on "File"<br>Click on "Save As" |  |
| **77.** | Save the guilford.txt file as<br>**brahe.txt** |  |
| **78.** | Close all windows |  |

<table>
<tr><td>79.</td><td colspan="2">G. The Revised Guilford Analysis: Brahe.txt</td></tr>
<tr><td>80.</td><td>Let’s launch Facets again</td><td></td></tr>
<tr><td>81.</td><td>Click on “Files”<br>Click on “Specification File Name?”</td><td>
</td></tr>
<tr><td>82.</td><td>Click on “brahe.txt” and “Open”<br>or Double-Click on “brahe.txt”<br>“Extra Specifications” - click on “OK”<br>“What is the Report Output file name” - click on “Open”</td><td>
</td></tr>
<tr><td>83.</td><td>Table 1 in the Facets analysis window: notice that each of our 3 models has 35 responses assigned to it. Each judge rated 7 examinees on 5 items = 35 ratings.<br>Correct so far!</td><td>
</td></tr>
<tr><td>84.</td><td>Let’s produce a customized version of Table 6, the vertical “rulers”<br>On the Facets menu bar, click on “Output Tables”<br>Click on “Table 6: Vertical Rulers”</td><td>
</td></tr>
<tr><td>85.</td><td>Specify for the contents of the rulers, Vertical=<br>1A, 2A, 3A<br>This displays the alphabetical element labels for all three facets.<br><br>Click on “Temporary Output File”</td><td>
</td></tr>
</table>

**86.** Compare Table 6 with Table 6 in #44. There have been some changes.
Red box: Brahe is now shown as the most lenient judge.

Bottom red box: Two rating scale structures are shown to the right. S.1 and S.2. They are identified below the Figure.

Notice that the range of S.1 is 1 to 9, and of S.2 is 2 to 8. This is because Brahe did not use the extreme categories. “Keep” only keeps the intermediate unobserved categories, not the extreme ones.

If you want to keep unobserved extreme categories, then add “dummy” data records to your data file which include those categories.

Edward is now the most creative, by far.

```
Table 6.0  All Facet Vertical "Rulers".

Vertical = (1A,2A,3A,S) Yardstick (columns lines low high extreme)= 0,10,-1,2,End
+---------------------------------------------------------------------------------+
|Measr|+Senior scientists|+Junior Scientists|+Traits                  | S.1 | S.2 |
|-----+------------------+------------------+-------------------------+-----+-----|
|   2 +                  +                  +                         + (9) + (8) |
|     |                  |                  |                         |     |     |
|     |                  |                  |                         |  8  |     |
|     |                  |                  |                         |     |     |
|     |                  |                  |                         |     |  7  |
|     |                  |                  |                         |     |     |
|     |                  | Edward           |                         | --- |     |
|     |                  |                  |                         |     |     |
|   1 +                  +                  +                         +     +     |
|     |                  |                  |                         |  7  | --- |
|     |                  |                  |                         |     |     |
|     |                  |                  |                         |     |     |
|     |                  |                  |                         |     |  6  |
|     |                  |                  |                         | --- |     |
|     |                  | George           |                         |     |     |
|     | Brahe            |                  | Attack                  |  6  | --- |
|     |                  | Betty            |                         |     |     |
|     |                  |                  | Daring                  | --- |     |
*   0 * Cavendish        *                  * Basis                   *     *  5  *
|     | Avogadro         |                  |                         |  5  |     |
|     |                  | Anne             | Clarity     Enthusiasm  |     |     |
|     |                  |                  |                         | --- | --- |
|     |                  | Chris            |                         |     |     |
|     |                  | Fred             |                         |  4  |     |
|     |                  |                  |                         |     |  4  |
|     |                  | David            |                         | --- |     |
|     |                  |                  |                         |     |     |
|     |                  |                  |                         |  3  |     |
|  -1 +                  +                  +                         + (1) + (2) |
|-----+------------------+------------------+-------------------------+-----+-----|
|Measr|+Senior scientists|+Junior Scientists|+Traits                  | S.1 | S.2 |
+---------------------------------------------------------------------------------+

S.1: Model = 3,?,?,FORWARD  ; Senior scientists: Cavendish
S.2: Model = 2,?,?,REVERSED  ; Senior scientists: Brahe
```

**87.** On the Windows *Task bar (bottom of your screen)*, click on “brahe.out.txt”
Scroll down to Table 7.1.1.
Red box: Brahe (Outfit MnSq = 1.66) is still noticeably the worst fitting judge according to the Outfit statistics, but now Cavendish (Outfit MnSq = .46) is shown as overfitting: mean-square well below 1.0

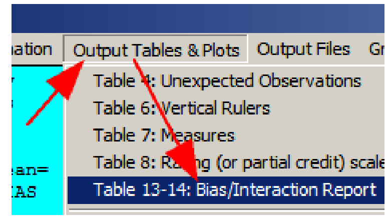
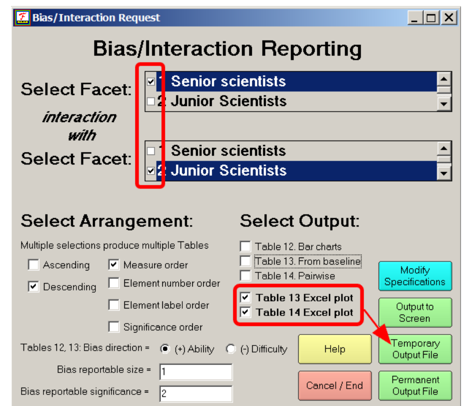
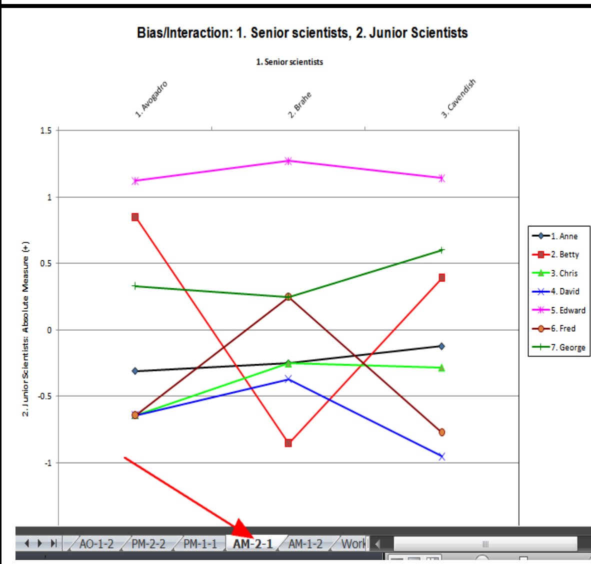
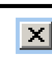
```
------------------------------------------------------------------------------
| Infit      Outfit    |Estim.| Correlation | Exact Agree. |
| MnSq ZStd  MnSq ZStd|Discrm| PtMea PtExp | Obs %  Exp % | N Senior sc
+----------+----------+------+-------------+-------------+------------------
| 1.67  2.5  1.66  2.4|  .17 |   .26   .60 |  14.3   20.5 | 2 Brahe
|  .46 -2.9   .46 -2.9| 1.62 |   .82   .67 |  32.9   28.0 | 3 Cavendish
|  .85  -.6   .87  -.5| 1.01 |   .78   .68 |  32.9   28.4 | 1 Avogadro
```

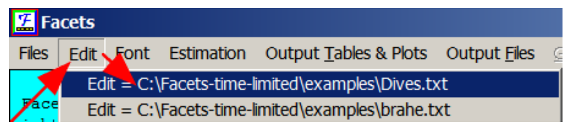
**88.** But *remember the suggested guideline:* **Remedy high mean-squares before low mean-squares!**
Red box: It is Brahe’s large noisy misfit that has forced Cavendish to appear to overfit. Remove Brahe, and Cavendish would fit reasonably well.

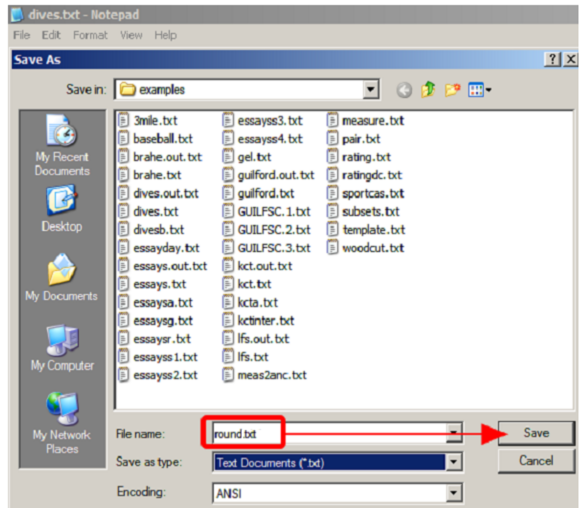
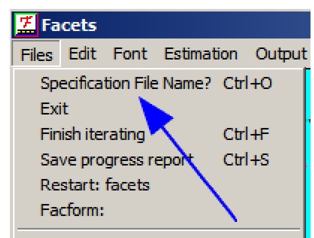
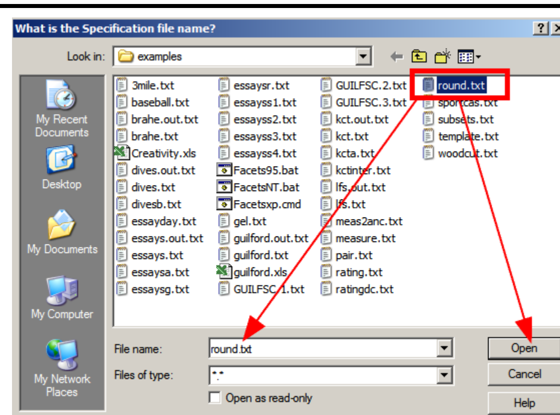
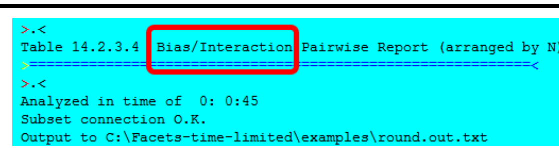
From an analysis without Brahe:

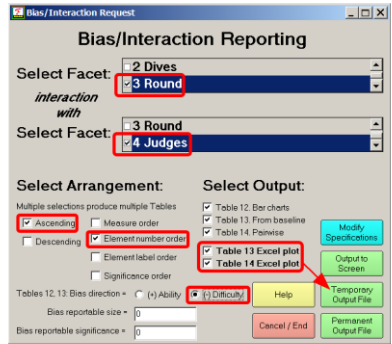
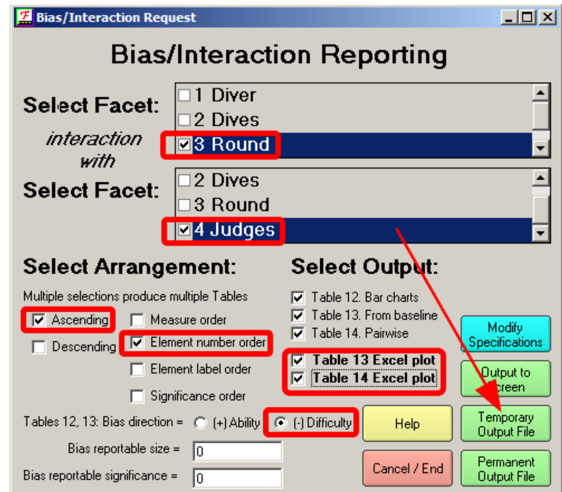
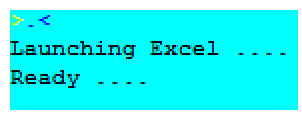
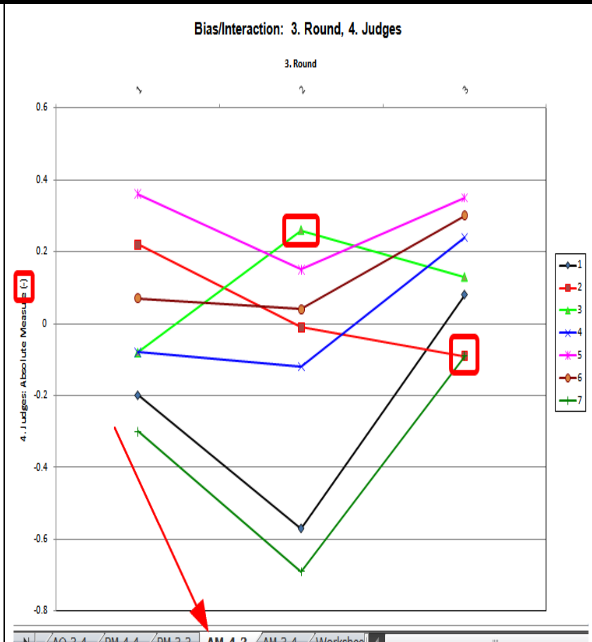
```
-------------------------------------------------------------------
| Infit      Outfit    |Estim.| Exact Agree. |                      |
| MnSq ZStd  MnSq ZStd|Discrm| Obs %  Exp % | N Senior scientists  |
-------------------------------------------------------------------
|  .99   .0   .98   .0|   .94 |  51.4   47.3 | 1 Avogadro          |
|  .73 -1.1   .76  -.9|  1.29 |  51.4   47.3 | 3 Cavendish         |
-------------------------------------------------------------------
```

**89.** Fit to a Rasch-measurement model is based on the ideal of *local independence* which says that the elements are "statistically independent after accounting for the main effects". The main effects, in this example, are the measures of rater severity, item difficulty and person ability.
We can only investigate *local independence* in the context of the current dataset, so raters who are locally independent in our dataset might not be locally independent if we change the data set.
*Facets* bases its fit statistics on the **average fit** of the elements to the model, so removing a worse fitting rater (Brahe) forces *Facets* to report the other two raters as fitting the Rasch model better. **Fit analysis is relative!**

<table>
<tr><td>90.</td><td colspan="2">H. The Revised Guilford Analysis: Brahe.txt: Interaction/Bias</td></tr>
<tr><td>91.</td><td>Do you wonder what happened on the bias/interaction plot? Let's take a look ....<br><br>On the main Facets menu,<br>Click on "Output Tables & Plots".<br>Click on "Table 13-14: Bias/Interaction Report"</td><td>
</td></tr>
<tr><td>92.</td><td>In the "Bias/Interaction Request" dialog box, check the two facets: 1. Senior scientists, 2. junior scientists<br>Then check "Table 13 Excel plot".<br>Then click on "Temporary Output File"<br><br>Tables 13 and 14 will be written to the temporary output file, but that is not what we want right now.<br><br>Excel will also be launched ..... It may take a few seconds to display ..</td><td>
</td></tr>
<tr><td>93.</td><td>When the Excel plot displays,<br>Click on tab: AM-2-1<br><br>Compare this to the same plot in #65. There is much more agreement, but the problem isn't solved.<br>In #65, the judges agreed that Betty was the most creative overall.<br>Now they agree that Edward is the most creative.<br><br>The Examination Board will have to decide: which analysis do we believe? Rasch measurement has done all that it can ....<br>Measurement and statistics can point out where the problems are, but human decision-making is still required!</td><td>
</td></tr>
<tr><td>94.</td><td>Close all windows</td><td></td></tr>
</table>

| 95. | **I. Interactions using Dummy Facets: Diving** | |
|---|---|---|
| **96.** | Do you remember the High-School Diving data in Dives.txt ? Each diver performed 3 dives, one per round. Here is a question: “Did the judges maintain their severity across the three rounds?” To find out, we need to investigate *judge x round* interactions. | The diving data:<br>diver, dive, round, judge, rating:<br>12, 2, 3, 7 ,11 |
| **97.** | But we don’t want to “round” to change the measurements of the other elements. We don’t want to include “round” as a main measurement effect. We only want “round” to investigate interactions. To do this, we will specify that “round” is a **dummy** facet that does not contribute to measurement, but is available for interactions. | *Measurement model:*<br>diver + dive + judge → rating<br><br>*Interaction model:*<br>judge x round → rating residual |
| **98.** | Launch *Facets* | |
| **99.** | Click on “Edit”<br>Click on “Dives.txt” (it will be somewhere on the list) |  |
| **100.** | In the *NotePad* window for Dives.txt,<br><br>Scroll down to “Round”, Facet 3 | `labels=`<br>`...`<br>`*`<br>`3,Round`<br>`1-3`<br>`*`<br>`4,Judges`<br>`......` |
| **101.** | To make elements available in a way that does not alter the other measures, we anchor (fix) their measures at 0.0. Then combining the elements of the anchored facet with elements of the other facets makes no difference to the measures. It is a “Dummy” facet.<br>To specify that a facet is anchored by putting an “A” (for “Anchored”) after the facet label. We specify the element anchor measures are 0 after the element labels .<br>Or we can put a “D” after the facet label, then there is no need to specify the anchor values. | Dummy facet:<br><br>`3,Round, A`<br>`1-3, , 0`<br>`*`<br>or<br><br>`3,Round, D`<br>`1-3`<br>`*` |
| **102.** | So please change Dives.txt<br>Since the “Round” elements are anchored, they won’t be reported during estimation. | `3,Round, A`<br>`1-3, , 0`<br>`*` |

| | | |
|---|---|---|
| **103.** | We also need to change the *Model=* specification to<br>1. activate the facet, and<br>2. specify the bias/interaction analysis:<br>This was:<br>`Models =`<br>`7,8,?,?,M`<br>`?,?,X,?,DoublePoints`<br>`*`<br>The "Round" facet is ignored for measurement "X" | Change to:<br>`Models =`<br>`7,8,?,?,M`<br>`?,?,?B,?B,DoublePoints`<br>`*`<br>The "Round" facet is now active "?". But since its elements are anchored at 0, it won't alter the measurements. Interactions between facet 3 (round) and facet 4 (judges) will be computed, "?B,?B" |
| **104.** | After you have made the changes,<br>*NotePad* Edit window:<br>Click on "File"<br>Click on "Save as"<br>Type in "**round.txt**"<br>Click on "Save" |  |
| **105.** | On the Facets Menu bar<br>Click on "Files"<br>Click on "Specification File Name?" |  |
| **106.** | Click on "**round.txt**" and "Open"<br>or Double-Click on "round.txt"<br>"Extra Specifications" - click on "OK"<br>"What is the Report Output file name" - click on "Open" |  |
| **107.** | In the Facets window, the iteration report ends with the Bias/Interaction analysis. |  |

| | | |
|---|---|---|
| **108.** | Let’s see if there are any interesting round-effects:<br>Facets menu bar<br>Click on “Output Tables & Plots”<br>Click on “Table 13-14: Bias/Interaction Report” |  |
| **109.** | “Bias/Interaction Request” dialog box<br>Check “3 Round”<br>Check “4 Judges”<br><br>Select Output:<br>Check “Table 13 Excel plot”<br>Check “Table 14 Excel plot”<br><br>For Tables 13 and 14:<br>Check “Ascending”<br>Check “Element number order”<br>Click on “Bias direction = Difficulty”<br><br>Click on “Temporary Output File” |  |
| **110.** | Tables 13 and 14 are shown almost immediately in a *NotePad* window.<br>Excel launches: |  |
| **111.** | Shortly afterwards, the Excel plot is displayed.<br>Click on the Excel tab: AM-4-3<br>The x-axis values (columns) are the Rounds.<br>The y-axis is Judge severity. The “(-)” in the y-axis title indicates that “higher measure = lower score”.<br><br>Again, this is interesting. There is a general trend among most judges.<br>“Bias direction = Difficulty”, so a lower y-value means “less difficult = less severe.” The Judges appear to become less severe (more lenient) in round 2, and then more severe in round 3.<br>*Warning!* This pattern could be due to changes in judge behavior across round, or systematic changes in diver performance across round (better performances in round 2). We don’t know. But let’s attribute it to judge behavior.<br>Red boxes: Notice the exceptions: Judge 3, the light green line, becomes more severe in round 2, and Judge 2, the red line, becomes more lenient in round 3. |  |

**112.** Let’s confirm our findings. Look at the NotePad window for Tables 13-14. It is the most recent Notepad window on the Windows Task bar. It is a temporary file so it has a cryptic name.


(Your Window name will differ.)

**113.** In Table 13, we see that for Judge 2 across the 3 rounds, his Observed-Expected average is increasing. He is giving increasingly higher than expected scores. His severity is decreasing.
But the significance of the changes, the *t*-statistics, are non-significant $(t < 2.0)$ because we only have a few observations by each judge in each round.

| Obs-Exp Average | Bias Size | Model S.E. | t | d.f. | Prob. | Infit MnSq | Outfit MnSq | Round Sq | Round N | Round R | Round measr | Judges N | Judges J | Judges measr |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| -.04 | .03 | .25 | .11 | 10 | .9111 | .9 | 1.0 | 1 | 1 | 1 | .00 | 1 | 1 | -.23 |
| .50 | -.33 | .24 | -1.40 | 11 | .1898 | .5 | .4 | 2 | 2 | 2 | .00 | 1 | 1 | -.23 |
| -.50 | .31 | .24 | 1.32 | 10 | .2156 | 1.2 | 1.2 | 3 | 3 | 3 | .00 | 1 | 1 | -.23 |
| -.28 | .18 | .24 | .76 | 10 | .4667 | 1.1 | 1.1 | 4 | 1 | 1 | .00 | 2 | 2 | .03 |
| .07 | -.05 | .23 | -.19 | 11 | .8492 | .8 | .8 | 5 | 2 | 2 | .00 | 2 | 2 | .03 |
| .20 | -.13 | .24 | -.52 | 10 | .6118 | 1.3 | 1.3 | 6 | 3 | 3 | .00 | 2 | 2 | .03 |

**114.** Look at the bottom of Table 13.
There is a chi-square test of the hypothesis: “The biases shown in this Table are all the same apart from measurement error”. The probability of this hypothesis is 0.96. So we certainly cannot reject the hypothesis of *no bias overall.* But the pattern of small interactions is interesting!
So here we truly have the familiar conclusion: *“More research is necessary!”*

When should we adjust for Differential Item Functioning? See *Optional Reading* at #136

| Obsvd Score | Exp. Score | Obsvd Count | Obs-Exp Average | Bias Size | Model S.E. | t | Infit MnSq | Outfit MnSq | Sq N Roun |
|---|---|---|---|---|---|---|---|---|---|
| 116.5 | 116.5 | 11.0 | -.01 | .00 | .24 | .00 | .9 | .9 | Mean (Cou |
| 11.6 | 10.6 | .9 | .27 | .18 | .01 | .73 | .3 | .3 | S.D. (Pop |
| 11.9 | 10.9 | .9 | .28 | .18 | .01 | .75 | .3 | .3 | S.D. (Sam |

Fixed (all = 0) chi-square: 11.2 d.f.: 21 significance (probability): .96

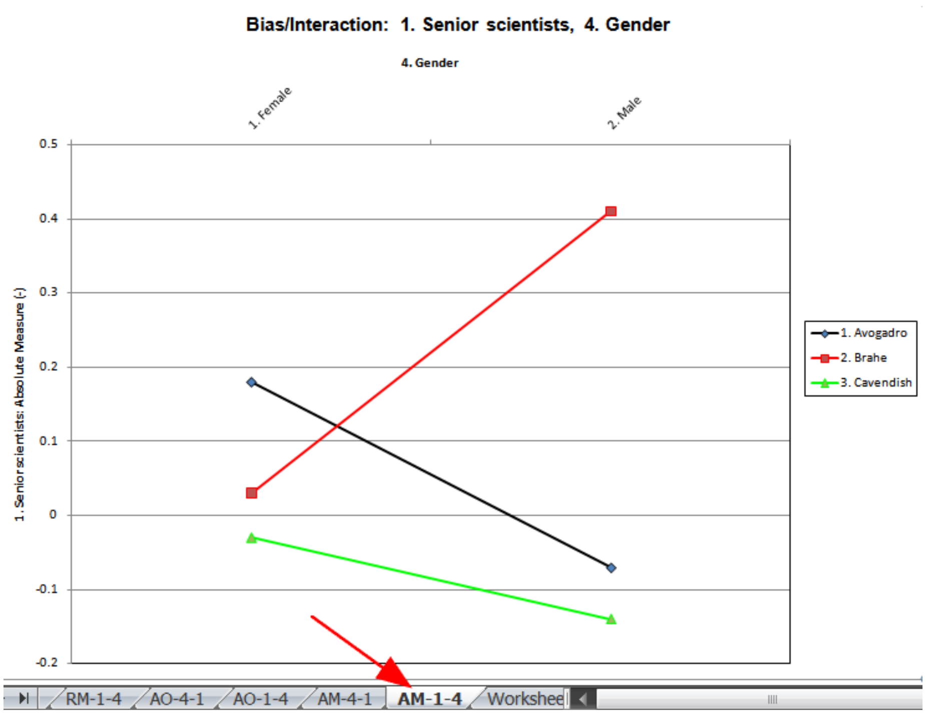
**115.** Close all windows

**116.**

| | | |
|---|---|---|
| **117.** | **J. Dummy Demographic Facets** | |
| **118.** | In the Dives data, the dummy facet was an independent facet (diving round), but often we want a dummy facet for interactions that are part of the person or item label. Here is the ***Knox Cube Test*** data. What if we want to investigate *Differential Item Functioning (DIF)* between Boy/Girl gender and the items? | Labels =<br>1,Children<br>1-17=Boy,,1<br>18-35=Girl,,2<br>*<br>2,Tapping items<br>1=1-4<br>2=2-3<br>3=1-2-4 |
| **119.** | 1. We increase the number of facets from 2 to 3.<br><br>2. We add another "?" to the Models= specification<br>We specify the interaction we want to investigate with "B" ... "B"<br><br>3. We add a dummy facet of Gender with two elements, Boy and Girl, both anchored, "A", at 0.<br><br>4A. We add facet 3, Gender, to the data with "D" for "Dummy". The elements of this facet are 1 (Male, Boys) and 2 (Female, Girls).<br>or<br>4B. We add facet 3, Gender, to the data with "A" for "Anchoring". The elements of this facet are 1 (Male, Boys) and 2 (Female, Girls) anchored at 0. | Facets = 3<br>Models = ?, ?B, ?B, D<br>1,Children<br>1-17=Boy,,1<br>18-35=Girl,,2<br>*<br>2,Tapping items<br>1=1-4<br>....<br>18=4-1-3-4-2-1-4<br>*<br>3, Gender, A ; dummy facet (or "D")<br>1 = Male, 0 ; anchored at 0<br>2 = Female, 0<br>* |
| | *Now there are two approaches:*<br><br>5A. We can insert the gender elements directly into the data as a dummy facet. The dummy facet is a new third facet with two elements. | Data=<br>; Boy 1 on item 1 with Gender 1 (Male) produces observation 1<br>1, 1, 1, 1<br><br>; Girl 18 1 on items 1-18 with Gender 2 (Female) produces observations 1,1,...,0<br>18,1-18,2,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0 |
| **120.** | or<br>5B. Approach 5A can be hard work, particularly when it requires reformatting the data. Here is an easier way.<br><br>The Gender information we need for dummy facet 3 is already included in the facet 1 element labels ("Boy" and "Girl". | Labels =<br>1,Children<br>1-17=Boy,,1<br>18-35=Girl,,2<br>*<br>2,Tapping items<br>1=1-4<br>2=2-3<br>3=1-2-4 |

| | | |
|---|---|---|
| **121.** | Let's use the facet 1 element labels with *Dvalues*=.<br><br>No change the Data= section.<br><br>Add the letters **B** and **G** to the element labels of the dummy "Gender" facet 3. These are the first letters of the element labels in the "Children" facet 1.<br><br>Now for the crucial specification,<br>**Dvalues=**<br>**3, 1, 1, 1 ; facet 3 is the gender**<br>*****<br>*Dvalues* = ; this defines element numbers **not** specified in the Data= section<br>3, **3** is the facet whose elements are to be specified in the data<br>1, the element identifiers for facet 3 are in the element labels come from facet **1**<br>1, the element identification for facet 3 starts in column **1** of the element 1 labels<br>1 the element identification is **1** column wide | Facets = 3<br>Models = ?, ?B, ?B, D<br>1,Children<br>1-17=**B**oy,,1<br>18-35=**G**irl,,2<br>*<br>2,Tapping items<br>1=1-4<br>....<br>18=4-1-3-4-2-1-4<br>*<br>3, Gender, A ; dummy facet (or "D")<br>1 = **B** Male, 0 ; anchored at 0<br>2 = **G** Female, 0<br>*<br>**Dvalues=**<br>**3, 1, 1, 1 ; facet 3 is the gender**<br>*****<br>Data =<br>1 ,1 ,1 ; Here are the original data<br>1 ,2-18,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0,0,0<br>2 ,1-18,1,1,1,1,1,1,0,0,1,1,1,0,0,1,0,0,0,0 |
| **122.** | *Here's how Dvalues= works:*<br><br>Instead of entering the element number of the demographic facet in the data file. *Dvalues*= tells us how to discover the demographic element number using the labels of the other facets.<br><br>This looks complicated, but it is fast and convenient - much easier than reformatting a data file to include a demographic variable.<br><br>For more examples and more options for *Dvalues=,* see *Facets* Help | **Data = 18, 3, 1** ; is one KCT observation<br>*Facets decodes this:*<br>18, facet 1 element 1 = child 18, "Girl"<br>3, facet 2 element 3 = item 3, "1-2-4"<br><br>**Dvalues=3, 1, 1, 1 ; facet 3 is the gender**<br>3, for facet 3<br>1, use the label for the facet 1 element.<br>The facet 1 element is 18, label: "Girl"<br>1, start at the first letter of label "**G**irl"<br>1 use 1 letter: "G"<br>Match "G" to the element labels of Facet 3.<br>"G" matches "G Female": element 2<br><br>*Facets analyzes* Data = 18, 3, 1 *as*<br>Data = **18, 3, 2, 1** |

| | | |
|---|---|---|
| **123.** | *More complicated example:*<br><br>Put in the element label all the demographic codes you want followed by the names, etc.<br><br>For instance: "Yoko" is Asian, Female, Teacher: | Labels=<br>Facets = 6 ; persons, items, raters, region, gender, occupation<br>1, persons<br>23 = AFT Yoko<br>...<br>*<br>4, Region, D ; this is a dummy facet for interactions, etc.<br>1 = Asia<br>2 = Europe<br>...<br>*<br>5, Gender, D<br>1 = Female<br>2 = Male<br>*<br>6, Occupation, D<br>1 = Student<br>2 = Teacher<br>3 = Administrator<br>*<br>dvalues =<br>4, 1, 1, 1 ; region code in facet 1 element label, 1st character, 1 column<br>5, 1, 2, 1 ; gender code in facet 1 element label, 2nd character, 1 column<br>6, 1, 3, 1 ; occupation code in facet 1 element label, 3rd character, 1 column<br>*<br>data=<br>23, 6, 15, 2 ; person 23 (Yoko), item 6, (region 2, gender 1, occ. 2,) rater 15, rating=2 |

| | |
|---|---|
| **124.** | **K. Interactions with Dummy Facets: Your Practical Task** |
| **125.** | You should have enough information to do all this yourself! So here is your task:<br>1. Open the Guilford.txt specification and data file in a *Notepad* Edit window (or any other way that you find convenient)<br>2. *Label:* the Junior Scientists with a gender: (Anne, Betty and Chris are female, the others are male).<br>3. *Label:* facet 4: gender<br>4. *Edit:* Facets=<br>5. *Edit:* Model= for 4 facets and "Senior Scientist x Gender" interactions<br>6. *Edit:* Zscore= 0,0 ; so that all interactions are reported in Table 14 etc.<br>7. *Add:* Dvalues=<br>8. Save your file under a new specification file name, such as Gender.txt<br>9. Analyze your new specification file with *Facets.*<br>10. Produce a plot of Differential Item Functioning: "Senior Scientist x Gender" |
| **126.** | ***Ready, Set, Go!*** |
| **127.** | |
| **128.** | Here's my plot does yours look anything like it? This plot shows the item difficulty (or gender performance) relative to their measures in the overall analysis.<br><br>If you want to see my specification file for this analysis it is example file: **G4.txt** |
| **129.** | Close all windows |

**130.** 

## L. Splitting Biased Items

**131.** What do we do about bias? If we decide that the item (or whatever) really is acting like two items, then split the item into two in the specifications and data.

In the KCT example, suppose it is item 8. We code two new items 8B for boys and 8G for girls, with missing data in the other part.

(In the RUMM2020 documentation, "splitting" an item is called "resolving" an item.)

```
2,Tapping items          ; Items are face
1=1-4                    ; Items labellec
2=2-3
3=1-2-4
4=1-3-4
5=2-1-4
6=3-4-1
7=1-4-3-2
; 8=1-4-2-3  ; commented out, so this
9=1-3-2-4
10=2-4-3-1
11=1-3-1-2-4
12=1-3-2-4-3
13=1-4-3-2-4
14=1-4-2-3-4-1
15=1-3-2-4-1-3
16=1-4-2-3-1-4
17=1-4-3-1-2-4
18=4-1-3-4-2-1-4
19=8B 1-4-2-3      ; new boy item
20=8G 1-4-2-3      ; new girl item
```

**132.** **Optional: Only do this if you want to!**

Rearrange the data: we need to change the number of items, and add the two half-columns of responses.

In Excel, copy and edit columns of responses.

In Word, alt+Mouse can be used for rectangular copy-and-pastes.

My version of this *Facets* specification file is at: kct-gender-8.txt

```
Data =                          ; no data file name, so data follows immediatel
1 ,1   ,1                       ; child 1 on item 1 scored 1
1 ,2-19,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0,0,0      ,0    ; item 19=8B
2 ,1-19,1,1,1,1,1,1,0,0,1,1,1,0,0,1,0,0,0,0    ,0
3 ,1-19,1,1,1,1,1,1,1,1,1,0,1,0,0,0,0,0,0,0    ,1
4 ,1-19,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0    ,1
5 ,1-19,1,1,1,1,1,1,1,1,1,1,0,0,1,1,0,0,0,0    ,1
6 ,1-19,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0    ,1
7 ,1-19,1,1,1,1,0,0,1,0,1,0,0,0,0,0,0,0,0,0    ,0
8 ,1-19,1,1,1,1,1,0,1,0,1,1,0,0,0,0,0,0,0,0    ,0
9 ,1-19,1,1,1,1,1,1,1,1,1,0,0,1,0,0,0,0,0,0    ,1
10,1-19,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0    ,1
11,1-19,1,1,1,1,1,1,1,1,1,0,0,1,0,0,0,0,0,0    ,1
12,1-19,1,1,1,0,1,1,0,0,0,0,0,0,0,0,0,0,0,0    ,0
13,1-19,1,1,1,1,1,1,1,1,0,1,0,0,0,0,0,0,0,0    ,1
14,1-19,1,1,1,1,1,1,1,1,1,0,1,0,1,0,0,0,0,0    ,1
15,1-19,1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0    ,1
16,1-19,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0    ,0
17,1-19,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0    ,1
18,1-20,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0    ,  ,1   ; item 20=8G
19,1-20,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0    ,  ,1
20,1-20,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0    ,  ,1
21,1-20,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0    ,  ,1
22,1-20,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0    ,  ,1
23,1-20,1,1,1,1,0,0,1,0,0,1,0,0,0,0,0,0,0,0    ,  ,0
24,1-20,1,1,1,1,1,0,0,1,1,1,1,1,0,0,0,0,0,0    ,  ,1
25,1-20,1,1,1,0,1,1,1,1,1,0,0,0,0,0,0,0,0,0    ,  ,1
26,1-20,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0,0,0    ,  ,0
27,1-20,1,1,1,1,1,1,1,1,1,1,1,0,1,0,0,1,1,0    ,  ,1
28,1-20,1,1,1,1,0,1,1,1,1,1,0,0,0,0,0,0,0,0    ,  ,1
29,1-20,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0    ,  ,1
30,1-20,1,1,1,1,1,1,1,1,1,1,1,1,1,0,1,0,0,0    ,  ,1
31,1-20,1,1,1,1,1,1,1,1,1,1,0,1,0,0,0,0,0,0    ,  ,1
32,1-20,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0    ,  ,1
33,1-20,1,1,1,1,1,1,1,1,0,0,1,0,0,0,0,0,0,0    ,  ,1
34,1-20,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0    ,  ,1
35,1-20,1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0    ,  ,1
```

**133.** And the resulting item difficulties are in Table 7.

The difference here is 2.2 logits, a little larger than in the bias/interaction report above. This is because splitting the item has also caused all the children's measures to be revised.

| Obsvd Score | Obsvd Count | Obsvd Average | Fair-M Avrage | Measure | Model S.E. | Infit MnSq | Infit ZStd | Outfit MnSq | Outfit ZStd | Estim. Discrm | Nu Tapping items |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 34 | .0 | .00 | ( 6.33 | 1.86) | Maximum | | | | | 18 4-1-3-4-2-1-4 |
| 1 | 34 | .0 | .01 | 5.00 | 1.08 | .76 | .0 | .11 | 1.4 | 1.18 | 15 1-3-2-4-1-3 |
| 1 | 34 | .0 | .01 | 5.00 | 1.08 | .76 | .0 | .11 | 1.4 | 1.18 | 16 1-4-2-3-1-4 |
| 1 | 34 | .0 | .01 | 5.00 | 1.08 | .76 | .0 | .11 | 1.4 | 1.18 | 17 1-4-3-1-2-4 |
| 3 | 34 | .1 | .03 | 3.56 | .70 | 1.57 | 1.2 | 1.35 | .9 | .62 | 14 1-4-2-3-4-1 |
| 6 | 34 | .2 | .08 | 2.42 | .55 | 1.18 | .6 | 1.14 | .5 | .80 | 12 1-3-2-4-3 |
| 7 | 34 | .2 | .10 | 2.13 | .52 | .69 | -1.0 | .37 | -.2 | 1.35 | 13 1-4-3-2-4 |
| 12 | 34 | .4 | .27 | .97 | .45 | 1.07 | .3 | .76 | -.1 | .98 | 11 1-3-1-2-4 |
| 11 | 16 | .7 | .74 | -1.09 | .75 | .55 | -.9 | .36 | -.7 | 1.41 | 19 8B 1-4-2-3 |
| 24 | 34 | .7 | .80 | -1.44 | .49 | 1.07 | .3 | .89 | .1 | .94 | 10 2-4-3-1 |
| 30 | 34 | .9 | .96 | -3.27 | .64 | 1.20 | .6 | .95 | .6 | .84 | 6 3-4-1 |
| 30 | 34 | .9 | .96 | -3.27 | .64 | .60 | -1.1 | .21 | .0 | 1.35 | 9 1-3-2-4 |
| 16 | 18 | .9 | .96 | -3.27 | .86 | .46 | -1.0 | .18 | .6 | 1.42 | 20 8G 1-4-2-3 |
| 31 | 34 | .9 | .98 | -3.72 | .71 | 1.02 | .1 | .48 | .5 | 1.03 | 5 2-1-4 |
| 31 | 34 | .9 | .98 | -3.72 | .71 | 1.35 | .9 | 2.43 | 1.2 | .50 | 7 1-4-3-2 |
| 32 | 34 | .9 | .99 | -4.30 | .81 | .90 | .0 | .31 | .8 | 1.10 | 4 1-3-4 |

**134.**

**135.**

**136.** *Optional Reading:*
#114 - "When to adjust for Differential Item Functioning" - http://www.rasch.org/rmt/rmt91e.htm
"DIF matters: A practical approach to test if Differential Item Functioning makes a difference"- http://www.rasch.org/rmt/rmt204d.htm
For an excellent description and summary of MFRM: "Many-Facet Rasch Measurement" by Thomas Eckes - http://www.coe.int/t/dg4/linguistic/Source/CEF-refSupp-SectionH.pdf

**137.**