Multiple Link Forms

If we can equate two forms, or one form to a bank, we can also equate multiple forms with multiple interconnections. We can use this same analysis to proceed one link at a time until eventually the entire network is equated. Any redundancies can be used to monitor and control the process. For example, linking form *A* to form *C* should give the same result as linking form *A* to form *B* to form *C*. Or, alternatively, linking *A to B to C to A* should bring us back to where we started and result in a zero shift, within statistical limits.

There is a straightforward least squares path to resolving any inconsistencies due to random noise. If we have *k* forms and there is a link t_{ij} between each pair of forms *(i, j)*, then summing over all the links for form *i*,

26. $$T_i = \sum_{j=1}^{k} t_{ij} = \sum_{j=1}^{k} (t_i - t_j) = kt_i - \sum_{j=1}^{k} t_j \ ,$$

where T_i is the sum of all form *i* links, t_{ij} is the link for form *i* to form *j*, and t_i is the general equating constant for form *i* that we are after. Then if all form-to-form links are present and if we adopt the convention $\sum_{j=1}^{k} t_j = 0$, the equating constant for form *i* is simply the row average:

27. $$t_i = \frac{T_i}{k} \ .$$

If not all the links are present, the solution is a little more complicated involving at worst a system of *k* simultaneous equations. In matrix form,

28. $$\boldsymbol{A}\underline{\boldsymbol{t}} = \underline{\boldsymbol{T}},$$

where $\underline{\boldsymbol{T}}$ is a *kx1* vector of the row sums T_i from equation (31), and $\underline{\boldsymbol{t}}$ is the vector of equating constants we are after. If all *k* links are present, ***A*** is a diagonal matrix with the value *k* along on the diagonal and expression (32) is the solution to expression (33).

In general, however, ***A*** is symmetric with:

- Diagonal elements $a_{ii} = k - m_i$, where *k* is the number of forms and m_i is the number of missing links for form *i*, and
- Off-diagonal elements $a_{ij} = 0$ if the link *(i,j)* is present and 1 if not, for $i \neq j$.

This tactic of *completing the sum* should seem vaguely familiar from the discussion of estimating the item difficulties. It can be used almost anywhere involving paired comparisons, like items or forms or football teams.

To illustrate, assume we have five forms with the form-to-form equating constants shown in table 10. The values in the table are <u>added</u> to the *row form* to equate it to the *column form*. Equating *A* to *B* means adding *-0.49* to form *A* logits; alternatively, equating *B* to *A* means adding *0.49* to form *B* logits. The value in the final column is added to the row form to equate it to the Bank. The first section of the table is completely filled so the equating constant for each form is the row mean.

Table 10: Resolution of Multiple Links

	Completely Filled Table of Links						Equating Constant
	A	B	C	D	E	Sum	
A	0.0	-0.49	-1.10	-1.47	-1.94	-5.00	-1.00
B	0.49	0.0	-0.41	-0.88	-1.71	-2.50	-0.51
C	1.10	0.41	0.0	-0.42	-0.90	0.19	0.04
D	1.47	0.88	0.42	0.0	-0.64	2.13	0.43
E	1.94	1.71	0.90	0.64	0.0	5.18	1.04
	Missing Links						
A	0.0	-0.49				-0.49	-0.93
B	0.49	0.0	-0.41			0.08	-0.44
C		0.41	0.0	-0.42		-0.01	-0.03
D			0.42	0.0	-0.64	-0.22	0.39
E				0.64	0.0	0.64	1.03

The second half of the table is missing six links: A-C, A-D, A-E, B-D, B-E and C-E (and hence 12 empty cells.) Finding the equating constants requires solving the matrix equation defined by expressions (32) and (33):

29.

$$\begin{bmatrix} 2 & 0 & 1 & 1 & 1 \\ 0 & 3 & 0 & 1 & 1 \\ 1 & 0 & 3 & 0 & 1 \\ 1 & 1 & 0 & 3 & 0 \\ 1 & 1 & 1 & 0 & 2 \end{bmatrix} * \begin{bmatrix} t_A \\ t_B \\ t_C \\ t_D \\ t_E \end{bmatrix} = \begin{bmatrix} -.49 \\ 0.08 \\ -.01 \\ -.22 \\ 0.64 \end{bmatrix}.$$

The two solutions give slightly different answers, but the second is missing a lot of data. In either case, adding the equating constant to every logit for a form will shift it to the common origin, centered on the mean of all forms. The only hard part is remembering when to add and when to subtract.

Controlling the non-Missing Links

The first half of Table 10 has a lot of redundancies that should be used to monitor and control the process. The equating of forms *A* and *C*, for example, is estimated directly by the observation $t_{ac} = -1.10$, which means *A* is 1.1 logits easier than *C*. It can also be estimated less directly by $t_{ab} + t_{bc} = -0.49 - 0.41 = -0.90$, which means *A* is *0.49* logits easier than *B* and *B 0.41* logits easier than *C*. Whether *0.9* is statistically different from 1.1 depends on the standard errors of the t_{ij}.

We could also approach the problems using the estimates from expression 33:

30. $\hat{t}_{ij} = t_i - t_j$.

Expression 35 was used to complete the entire matrix of improved estimates, which are shown in Table 11. These values should compare to Table 10 and any anomalies investigated.

Table 11: Fitted Values of Multiple Links

	Completely Filled Table of Links						Equating Constant
Form	A	B	C	D	E	Sum	
A	0.00	-0.49	-1.04	-1.43	-2.04	-5.00	-1.00
B	0.49	0.00	-0.55	-0.94	-1.55	-2.55	-0.51
C	1.04	0.55	0.00	-0.39	-1.00	0.20	0.04
D	1.43	0.94	0.39	0.00	-0.61	2.15	0.43
E	2.04	1.55	1.00	0.61	0.00	5.20	1.04

Continuing the A-B-C example with the improved values, t_{ac} equals *-1.04 = -0.49 – 0.55*, no matter how we get there. The relevant form to form links, estimated from the margins, are: *A to C is -1.0 - 0.04 = -1.04, A to B is -1.0 + 0.51 = -0.49,* and *B to C is -0.51- 0.04 = 0.55*. No matter how circuitous the route we follow from A to C, it will always be -1.04 logits and from C to A will always be 1.04. Unfortunately, the standard errors are cumulative. For any path, the standard error is the sum of the standard errors for each link in the path, which are based on the standard errors of estimation for each item difficulty in the link.

Identifying Confounded Connections

Rasch's original problem with oral reading arose because the data could not answer the basic question, *Were changes in the reading scores because the student had improved or because the test was different?* The data were not properly connected, or in the language of experimental design, the *occasion* effect was confounded with the *test* effect, or in the language of mathematics, the model was not adequately identified.

The specific objectivity of his new model allowed Rasch to construct the connections using entirely new samples. The design, shown in Table 11 (Rasch, 1960, p. 5), allowed Rasch to collect the new data needed to estimate the differences between successive texts, which in turn allowed him to separate the text effect from the occasion effect.

Table 11: Design of Remedial Reading Linking Study

	Test				
Grade	ORF	ORU	ORS	OR5	OR6
2	X				
3	X	X			
4			X	X	
5		X	X	X	
6			X	X	X
7				X	X

This sort of design is now common practice with either overlapping subtests or overlapping samples or both.

Linking forms does not impose any constraints on the items used other than they satisfy Rasch's principles. However, prudent link designs, both within and between forms, follow many of the same principles as good experimental design. The link items in a form should cover a range of difficulties and a balance of content and item types. In theory, this isn't necessary but *in theory, theory and practice are the same thing; in practice,* *they aren't*[1]. Covering a range of issues that might make items behave differently provides another opportunity to, in Rasch's terms, control the model. It can make it possible to eliminate threats to specific objectivity or, perhaps, gain some insights.

Unconventional Equating

Equating is such an important and defined task, treating it as a unique step in the item and bank development process is convenient, understandable, and perhaps desirable. The process we have just described uses the logit difficulties to equate two or more sets of calibrated items without reverting to the individual item responses. We have two sets of estimates for the same items with standard errors. Are the two sets different? This was the standard operating procedure for decades and has worked fine.

From a broader perspective, any calibration, even a fixed form on a single group, is an equating study. We are always trying to establish item difficulties that will yield equivalent logit scores for people regardless of which selection of items we choose to administer. The entire equating issue can be attacked using the machinery we have been developing in chapters *IV* and *V*. If everything conforms to Rasch's principles, you will get the same answers either way but that's a big if.

"Concurrent," "Simultaneous," and "Anchored" calibrations are a short step away from fixed form calibrations followed by equating studies like those just described. With a calibration method that allows missing data[2], (e.g., pairwise, chapter *III*) multiple forms and multiple administrations can be consolidated into a single *concurrent* calibration process, assuming they have something in common (i.e., they are linked/connected with common people or common items.) If we know (or think we know) the logit difficulties for some or all of the items, we can *anchor* those difficulties to the known values (rather than imposing our standard convention of $\Sigma d = 0$), which will produce estimates for the new items that are relative to the known values. If all the difficulties are known, this is just a check that they are still appropriate in the new occasion.

It has been a rather common practice in large scale assessment to use a fixed, operational form, composed of items that have been tried out and calibrated, that all students take, and that are the basis of the student measures. To this fixed form, multiple matrix forms of untried items are appended[3]. Thus each student takes the fixed form plus a few new items. The process can included as many matrix forms as necessary to try out all the new items needed to keep the item bank completely restocked and refreshed. There are several ways to do the arithmetic but the paired approach is naturally oblivious to the holes in this design because everything is connected through the operational form.

The equating problem is just another level of model control. We group the people by matrix form and use our favorite fit statistic to find the operational items that didn't work right in any forms. Or if we are equating this year's form to last year's, we split the

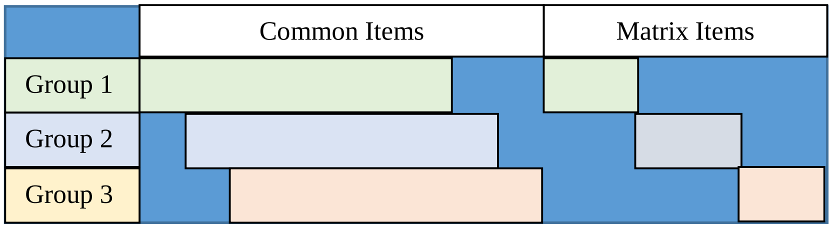
---

[1] Paraphrasing Yogi Berra a little.

[2] All the standard software packages can do this, but may make slightly different assumptions and so get slightly different answers when the data don't conform perfectly.

[3] Ideally, they are not just tacked on at the end, but are embedded throughout. This is big deal with printed forms, not so much with computer administered.

people between current and previous and examine the overlapping items. For Rasch, it's just another check on specific objectivity.



The diagram above attempts to suggest some possibilities for a link design. There are three groups of examinees, defined by the matrix items each took. Ideally, the groups will be statistically equivalent, by some random assignment process, but that is just good practice not a Rasch requirement. There is no overlap in items between the matrix forms (otherwise those items would be common.) They may be new and uncalibrated or not; they just do not provide a direct connection between groups.

The so-called common items do provide the connections between groups. They may be completely in common across all groups, as with the operation form scenario just described, or there may be varying degrees of overlap. Some groups may not be connected directly to every other group.

Exactly how one proceeds and what gets anchored to what depend on the specifics and motivations of the situation. For example, if we are equating this year's form to last year's, the "matrix" portion will relatively large, perhaps 60 out of 75 items. The concern here is to minimize the impact of the exposure of last year's items. The link then would be the 15 or so items that are in common between the two years. Alternatively, if our intent is try-out some completely new items by embedding them in an operational form, the numbers will be more or less reversed with the bulk in the common (i.e., operational, i.e., link) portion. The concern here is not to overburden any one examinee.

In either case, we always try not to let any faulty item influence the definition of the construct. For the operational-with-embedded-field-test design, this means calibrating the operational form on all examinee groups without including any field test items. When that has been completed, anchor the operational items and calibrate the field tests. When equating this year to last year, it means calibrating everything together.

The Pair matrix of counts should look something like the diagram below, implying every matrix form is connected to the common but no other matrix.

	Common	Matrix 1	Matrix 2	Matrix 3
Common	X	X	X	X
Matrix 1	X	X	0	0
Matrix 2	X	0	X	0
Matrix 3	X	0	0	X

As discussed earlier, this matrix of counts, as well as its derivatives of log odds and logits, provides many controls on the process, e.g., do the margins adequately recover the entries? Are there specific items or item types that stick out? But to have total control, we need to break the counts up by examinee group: Does specific objectivity hold across gender, ethnicity, administration, grade, age, etc.?

As always, objectivity is specific to the threats eliminated.