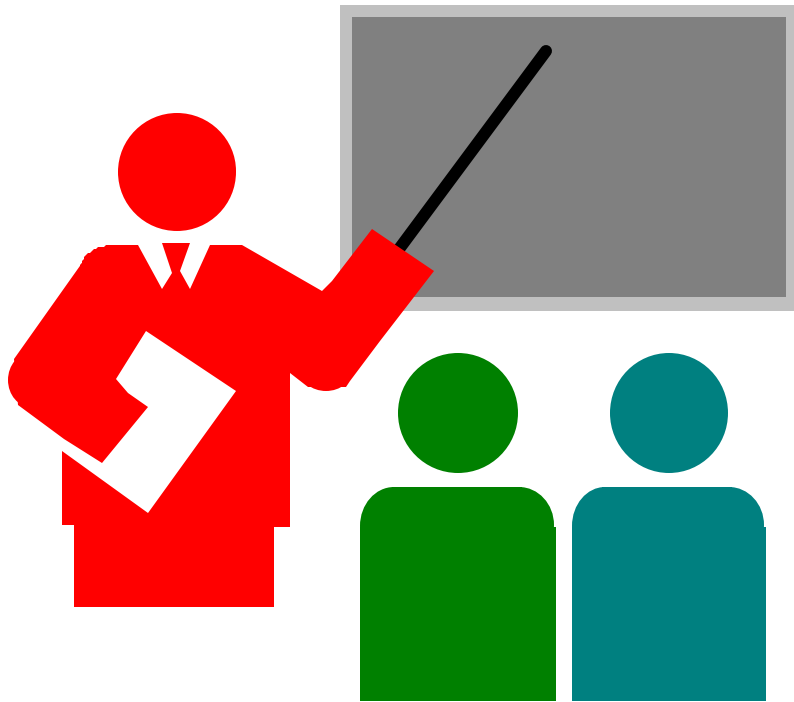


TEST EQUATING



- Tests typically don't have the same mean, standard deviation, and reliability coefficient.
- Puts scores from two or more tests on the same measurement scale so they can be compared.

Test Equating

Test Administration Results

Twenty Students took two 20 item math tests during the first six week grading period.

Test 1

Mean = 15.2

S.D. = 2.9

SEM = 1.7

Alpha = .68

Test 2

Mean = 16.4

S.D. = 2.9

SEM = 1.5

Alpha = .74

Rasch Calibration Program

Rasch Control File (BIGSTEPS)

```
;File TEST2.TXT
&INST
TITLE='Test 2 item analysis'
NI=20                                ;number of items
DATA=TEST2.DAT                      ;name of data file
ITEM1=1
NAME1=21
TABLES=11111111111111111111111111
PERSON=Person
ITEM=Math
STBIAS=Y                            ;adjust for UCON estimation bias
PFILE=TEST2.PF                      ;write person measures to a file
IFILE=TEST2.IF                      ;write item calibration to a file
&END
1=1                                ;first item name
.
.
20=20                              ;last item name
END NAMES
```

Rasch Equating Program

Rasch Anchor File (BIGSTEPS)

```
; This file is EQUATE1.TXT
&INST
TITLE='EQUATE TEST 1 TO TEST 2- 5 ANCHORED ITEMS'
NI=20
ITEM1=1
NAME1=21
PFILE=EQUATE1.PF
IFILE=EQUATE1.IF
PERSON=PERSON
ITEM=Math
DATA=TEST1.DAT ; name of data file (anchoring test 2 to test 1)
IAFILE=* ; list of items to be anchored (16 to 20)
16 -1.74 ; item 16 anchored at the measure of -1.74 ( item difficulties)
17 -2.06
18 -.63
19 1.68
20 1.40 ; item 20 anchored at the measure of 1.40 ( item difficulties)
*
&END
1-20 ;item names
END NAMES ; End of this file
```

Test Equating

(Test 1 Equated to Test 2)

TEST 1

PERSON	MEASURE	SCORE	ERROR
1	1.52	14	.59
2	2.30	16	.68
3	4.39	19	1.45
4	1.52	14	.59
5	1.18	13	.57
6	1.88	15	.62
7	3.64	18	1.05
8	.57	11	.55
9	1.88	15	.62
10	2.83	17	.79
11	.57	11	.55
12	.87	12	.55
13	4.39	19	1.45
14	4.39	19	1.45
15	-.02	9	.54
16	1.18	13	.57
17	1.52	14	.59
18	1.52	14	.59
19	1.88	15	.62
20	.28	10	.54

TEST 2

PERSON	MEASURE	SCORE	ERROR
1	1.16	13	.58
2	2.52	16	.78
3	3.36	17	1.04
4	-.06	9	.52
5	3.36	17	1.04
6	3.36	17	1.04
7	2.52	16	.78
8	1.97	15	.68
9	4.12	18	1.40
10	1.16	13	.58
11	2.52	16	.78
12	3.36	17	1.04
13	.76	11	.56
14	1.97	15	.68
15	2.50	15	.79
16	2.52	16	.78
17	2.52	16	.78
18	.44	10	.55
19	2.52	16	.78
20	-.17	8	.53

Test Equating

Score Scale Conversion

- ◆ Parents don't understand “logits” so we should convert logit measures to a scale score.
- ◆ A common score scale is the National Curve Equivalent
$$\text{NCE} = 50 + \text{logit} * 21.06$$
- ◆ Converting both sets of logit measures for students permits a NCE score comparison from 0 to 99.
- ◆ We can now answer the question:
 - On which test did a student do better?

Test Equating

- ◆ Common Items or Common Persons can be used
- ◆ Common Items should possess certain characteristics
 - clean - good wording and understandable
 - close - near the average ability level of students
 - consistent - don't vary from sample to sample
 - control - items are content valid
 - constant - narrow standard deviation of item difficulties
- ◆ Approximately 20 % of items should be used for anchoring