

Evaluating sufficiency of measurement: A judgmental approach

Chad Buckendahl
Buros Center for Testing

Purpose

- ▶ Reporting at multiple performance levels (e.g., basic, proficient, advanced)
- ▶ Millman & Greene (1989); Subkoviak (1984)
- ▶ Opportunities to demonstrate performance
 - Breadth – range of difficulty
 - Depth – number of opportunities

Methodology

- ▶ Create performance level descriptors
- ▶ Gather judgments from teachers on how well different tests measure desired levels of performance in reading and mathematics
 - Commercially available NRTs
 - District created CRTs

Some results - NRT

	<u>Adv.</u>	<u>Prof.</u>	<u>Prog.</u>	<u>Beg.</u>
▶ Reading				
▪ 4.1.3 (NRT-1)	5	15	12	0
▪ 4.1.3 (NRT-2)	9	17	11	3
▪ 4.1.3 (NRT-3)	5	23	10	5
▶ Mathematics				
▪ 8.5.2 (NRT-1)	1	6	6	1
▪ 8.5.2 (NRT-2)	1	2	3	0
▪ 8.5.2 (NRT-3)	1	6	3	0

Some results - CRT

	<u>Adv.</u>	<u>Prof.</u>	<u>Prog.</u>	<u>Beg.</u>
▶ Reading				
▪ 12.1.1 (1-E)	0	3	5	3
▪ 12.1.2 (1-E)	2	5	3	0
▪ 12.1.3 (1-E)	0	6	4	0
▶ Mathematics				
▪ 4.1.5 (1-W)	1	4	10	1
▪ 4.2.1 (1-W)	10	13	2	0
▪ 4.5.1 (1-W)	1	21	3	0

Conclusions

- ▶ Evidence of sufficiency impacts validity
 - Student performance classifications
 - Curriculum and instructional decisions
- ▶ Current NRTs/CRTs used by districts vary in their capacity to classify students at multiple performance levels
- ▶ Low confidence in classifications of students at multiple performance levels

Recommendations

- ▶ Evaluate your assessments' capacity to classify students into your program's performance categories
- ▶ Adopt performance level descriptors (PLDs) to guide development or revision
- ▶ Collapse data into two levels (met/not met) to have greater confidence in the classifications of performance

Full report is available at:

www.nde.state.ne.us/stars