Using Principles of Measurement to support score interpretation

Michael C. Rodriguez

May 15, 2015

Minnesota Assessment Group

2013-2014 Technical Manual for Minnesota's Title I and Title III Assessments

p. 80

Understanding Measurement Error
When interpreting test scores, it is important to remember that test scores contain some amount of measurement error. That is to say, test scores are not infallible measures of student characteristics.

Because measurement error tends to behave in a fairly random fashion, when aggregating over students, these errors in the measurement of students tend to cancel out. Chapter 8, “Reliability,” describes measures that provide evidence indicating measurement error on Minnesota assessments is within a tolerable range. Nevertheless, measurement error must always be considered when making score interpretations.

p. 81

Using Objective/Strand-Level Information

Strand or substrand level information can be useful as a preliminary survey to help identify skill areas in which further diagnosis is warranted. The standard error of measurement associated with these generally brief scales makes drawing inferences from them at the individual level very suspect; more confidence in inferences is gained when analyzing group averages. When considering data at the strand or substrand level, the error of measurement increases because the number of possible items is small. In order to provide comprehensive diagnostic data for each strand or substrand, the tests would have to be prohibitively lengthened. Once an area of possible weakness has been identified, supplementary data should be gathered to understand strengths and deficits.

p. 150

$$SEM = 12.1\sqrt{(1 - 0.88)} = 4.19 \quad (8.5)$$

Placing a one-SEM band around this scale score would result in a score range of 346 to 554 (that is, 350 ± 4.0). Furthermore, in the case of unbiased scores and if measurement error is normally distributed, then the *true scores* for approximately 68% of test takers would fall in the interval band created by adding and subtracting one SEM from their reported score. Thus, the chances are better than 2 out of 3 those students with an observed score of 350 and SEM = 4 would have an estimated true score within the interval 346–354. This interval is called a confidence interval or confidence band. By increasing the

p. 152

Measurement Error for Groups of Students

As is the case with individual student scores, district, school and classroom averages of scores are also influenced by measurement error. Averages, however, tend to be less affected by error than individual scores. Much of the error due to systematic factors (i.e., bias) can be avoided with a well-designed assessment instrument that is administered under appropriate and standardized conditions. The remaining random error present in any assessment cannot be fully eliminated, but for groups of students random error is apt to cancel out (i.e., average to zero). Some students score a little higher than their true score, while others score a little lower. The larger the number in the group, the more the canceling of errors tends to occur. The degree of confidence in the average score of a group is generally greater than for an individual score.

2013-2014 Yearbook Tables for Minnesota's Title I and Title III Assessments

2014 MCA-III Score Distribution

Grade 03 Reading

Scale Score	Online S.E.M.	Paper S.E.M.	Frequency	Percent	Cum. Frequency	Cum. Percent	Percentile Rank	Achievement Level
335	5.2	5.1	726	1.2	12464	20.0	19	D
336	5.1	5.1	775	1.2	13239	21.3	21	D
337	5.1	5.0	809	1.3	14048	22.6	22	D
338	5.0	5.0	770	1.2	14818	23.8	23	D
339	5.0	5.0	819	1.3	15637	25.1	24	D
340	5.0	5.0	848	1.4	16485	26.5	26	P
341	5.0	5.0	921	1.5	17406	27.9	27	P
342	5.0	5.0	967	1.6	18373	29.5	29	P
343	5.0	5.0	1001	1.6	19374	31.1	30	P
344	5.0	5.0	1023	1.6	20397	32.8	32	P
345	5.0	5.0	1066	1.7	21463	34.5	34	P
346	5.0	5.0	1118	1.8	22581	36.3	35	P
347	5.0	5.0	1152	1.8	23733	38.1	37	P
348	5.0	5.0	1170	1.9	24903	40.0	39	P
349	5.0	5.0	1189	1.9	26092	41.9	41	P
350	5.0	5.0	1313	2.1	27405	44.0	43	M
351	5.0	5.0	1212	1.9	28617	46.0	45	M
352	5.0	5.0	1251	2.0	29868	48.0	47	M
353	5.0	5.0	1255	2.0	31123	50.0	49	M
354	5.0	5.0	1290	2.1	32413	52.0	51	M
355	5.0	5.0	1284	2.1	33697	54.1	53	M
356	5.0	5.0	1329	2.1	35026	56.2	55	M
357	5.0	5.0	1247	2.0	36273	58.2	57	M
358	5.0	5.0	1281	2.1	37554	60.3	59	M
359	5.0	5.1	1357	2.2	38911	62.5	61	M

2014 MCA-III Score Distribution

Grade 08 Mathematics

Scale Score	Online S.E.M.	Paper S.E.M.	Frequency	Percent	Cum. Frequency	Cum. Percent	Percentile Rank	Achievement Level
847	3.1	3.1	1577	2.7	20574	34.9	34	P
848	3.1	3.0	1525	2.6	22099	37.4	36	P
849	3.1	3.0	1668	2.8	23767	40.3	39	P
850	3.1	3.0	1666	2.8	25433	43.1	42	M
851	3.0	3.0	1684	2.9	27117	45.9	45	M
852	3.0	3.0	1795	3.0	28912	49.0	47	M
853	3.0	3.1	1742	3.0	30654	51.9	50	M
854	3.0	3.1	1779	3.0	32433	55.0	53	M
855	3.0	3.1	1760	3.0	34193	57.9	56	M
856	3.0	3.2	1717	2.9	35910	60.8	59	M
857	3.0	3.2	1777	3.0	37687	63.9	62	M
858	3.0	3.2	1744	3.0	39431	66.8	65	M
859	2.9	3.3	1656	2.8	41087	69.6	68	M
860	2.9	3.3	1614	2.7	42701	72.4	71	M
861	2.9	3.4	1598	2.7	44299	75.1	74	E
862	3.0	3.4	1492	2.5	45791	77.6	76	E

p. 78

2014 MCA-III Subscale Correlations

Grade 08 Mathematics

Pearson Correlation Coefficients					
	Total Scale Score	Number & Operation	Algebra	Geometry & Measurement	Data Analysis & Probability
Total Scale Score	1.00	0.84	0.95	0.79	0.75
Number & Operation	0.84	1.00	0.75	0.62	0.57
Algebra	0.95	0.75	1.00	0.71	0.66
Geometry & Measurement	0.79	0.62	0.71	1.00	0.54
Data Analysis & Probability	0.75	0.57	0.66	0.54	1.00

p. 64

MCA-III Summary Statistics

Grade 08 Mathematics - Online

Total					
Scale	Range of Items	N	Mean	SD	Marginal Reliability
Total Scale Score	42	51469	851.80	13.70	0.93
Number & Operation Strand	6-8	51469	5.06	1.78	0.69
Algebra Strand	18-29	51469	5.19	1.96	0.84
Geometry & Measurement Strand	6-8	51469	5.17	1.68	0.61
Data Analysis & Probability Strand	6-7	51469	5.06	1.69	0.51

p. 160

2014 MCA-III Subscale Correlations

Grade 03 Reading

Pearson Correlation Coefficients			
	Total Scale Score	Literature	Information
Total Scale Score	1.00	0.94	0.93
Literature	0.94	1.00	0.80
Information	0.93	0.80	1.00

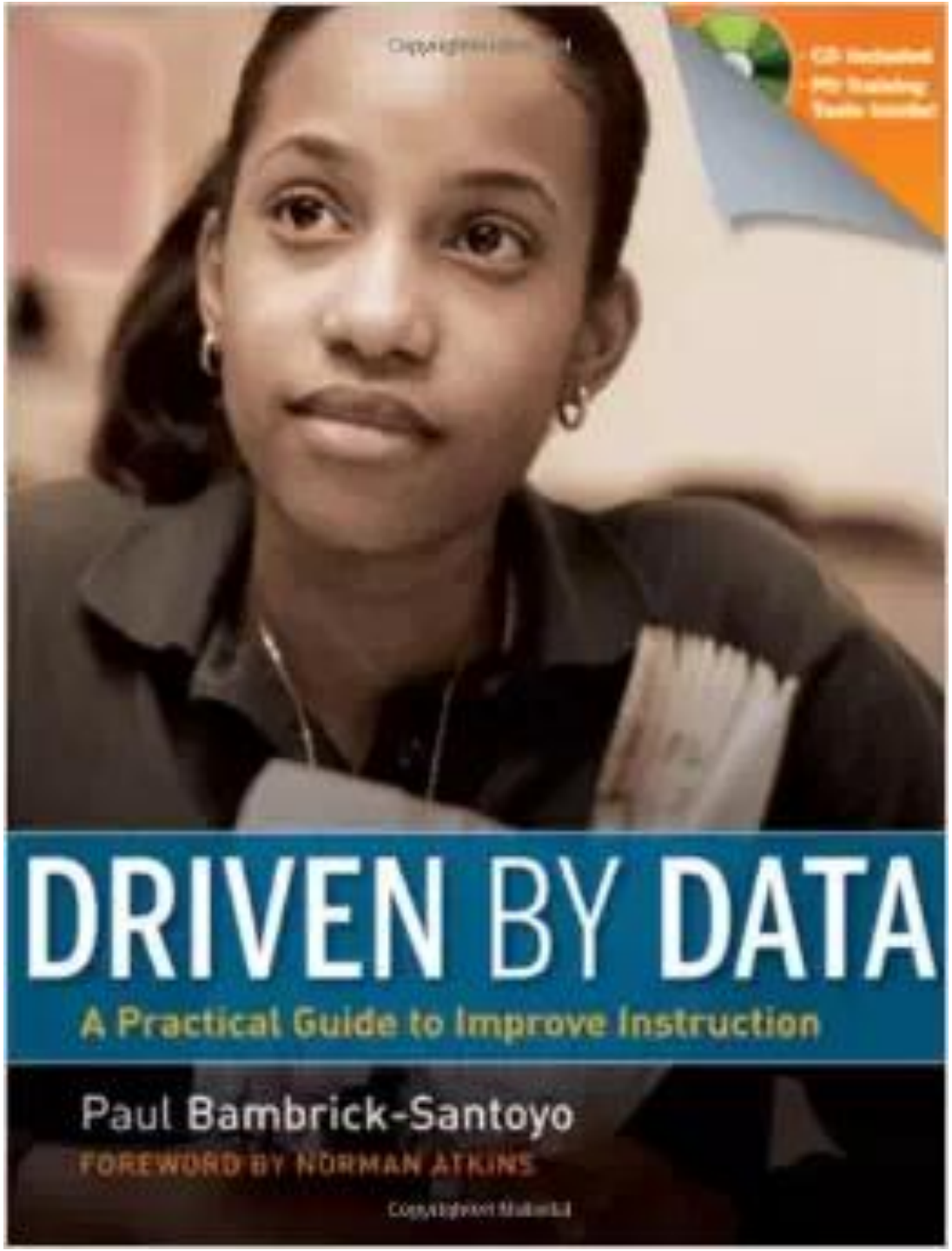
p. 133

2014 MCA-III Summary Statistics Reports

Grade 03 Reading - Online

Total					
Scale	Range of Items	N	Mean	SD	Marginal Reliability
Total Scale Score	48	24211	352.23	20.50	0.88
Literature	21-27	24211	5.04	1.90	0.81
Information	21-27	24211	5.06	1.94	0.80

$$r_{xy} \leq \sqrt{r_{xx} r_{yy}}$$


DRIVEN BY DATA
A Practical Guide to Improve Instruction
Paul Bambrick-Santoyo
FOREWORD BY NORMAN ATKINS

p. 8

CORE IDEA

- Assessments are not the end of the teaching and learning process; they're the starting point.

… we should not teach and then write an assessment to match; instead, we should create a rigorous and demanding test and then teach to meet its standards

p. 13

CORE IDEAS: Interim Assessments

- Start from the end-goal exam.
- Align the interim assessments to the end-goal test.

…

p. 28

Analyze the Interim Assessment or End-Goal Test

Acquire the closest version that you can find of your state test, interim assessment, or other year-end assessment by which your students' learning will be measured.

…