Using Flawed Data Creatively with JMP®:

...What To Do When Your Measurements Don't Measure Up

www.dibels.org

–Elizabeth N. Dewey, Senior Research Analyst / Statistician — Dynamic Measurement Group, Inc., Eugene, OR 97401

Introduction

In Psychometrics, relatively little research has been conducted on the effectiveness of individual test items that are chosen to reflect specific student skills. In our research towards the development of a new elementary mathematics assessment, a pilot study was conducted in first through fifth grades evaluating the efficiency of our new assessment, DIBELS® Math, in capturing student performance.

A total of 2,475 students were given two timed alternate forms of the same worksheet. Each worksheet contained 25 grade-level appropriate algebraic computation items organized within pre-determined categorical skill groups (e.g., addition, subtraction).

But when our worksheets didn't perform like we thought they should, we were forced to reevaluate our own methodology for item generation and inclusion. Instead of using the items to assess student skill, we decided to use student responses to assess item effectiveness.

Issues with the Data

Skipping and Missing Data

Students were asked to complete the worksheet in order, but skipping impacted the interpretability of the original analysis.

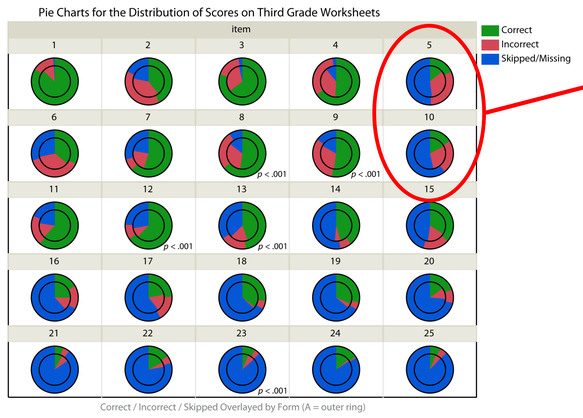
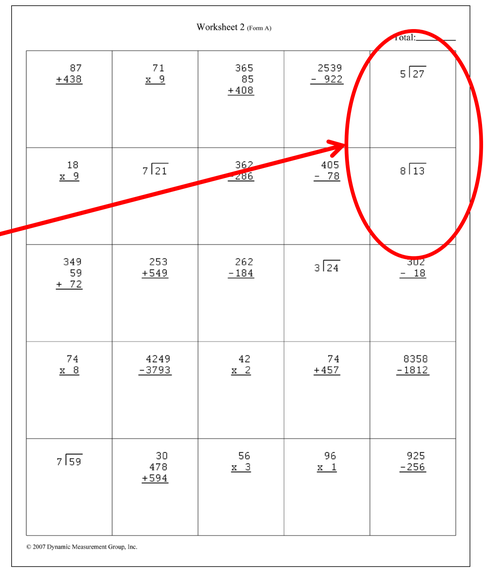
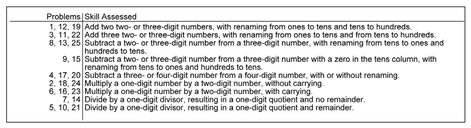


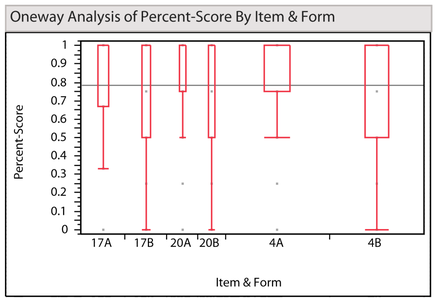
Figure 1. Pie charts organized like student worksheets illustrate patterns of skipping, which limited inference. Items 5 and 10 were often skipped in third grade.





Problems	Skill Assessed
1, 12, 19	Add two two- or three-digit numbers, with renaming from ones to tens and tens to hundreds.
3, 11, 22	Add three two- or three-digit numbers, with renaming from ones to tens and from tens to hundreds.
8, 13, 25	Subtract a two- or three-digit number from a three-digit number, with renaming from tens to ones and hundreds to tens.
9, 15	Subtract a two- or three-digit number from a three-digit number with a zero in the tens column, with renaming from tens to ones and hundreds to tens.
4, 17, 20	Subtract a three- or four-digit number from a four-digit number, with or without renaming.
2, 18, 24	Multiply a one-digit number by a two-digit number, without carrying.
6, 16, 23	Multiply a one-digit number by a two-digit number, with carrying.
7, 14	Divide by a one-digit divisor, resulting in a one-digit quotient and no remainder.
5, 10, 21	Divide by a one-digit divisor, resulting in a one-digit quotient and remainder.

Figure 2. The Third Grade worksheet and skill set chart.



1-way Test, ChiSquare Approximation

ChiSquare	DF	Prob>ChiSq	
17.6652	5	0.0034	*

Figure 3. Median test for differences within skill sets returns significant results.

Differences within Skill Sets

Students didn't attempt every item, thus evaluation of the items that appeared later were inconclusive. Additionally, significant differences between the items that we were able to evaluate suggested that our skill sets didn't function as intended. The question became:

What skills are we measuring, and how do we tell if our measurements are reliable?

We couldn't answer this question.

Methods of Exploration

Exploratory Factor Analysis (EFA) and Analysis of Means (ANOM)

The EFA components form a model that relates the skill sets to a latent immeasurable overall skill (algebraic computation) that influences responses on the test items (Child, 1990). Percent-scores were used for analysis, and factors were retained based on commonly used guidelines: positive eigenvalues, percentage of variance explained, a scree test, the size of the residuals, and interpretability (Nunnally & Bernstein, 1994, Kim & Mueller, 1978, Jolliffe, 2002).

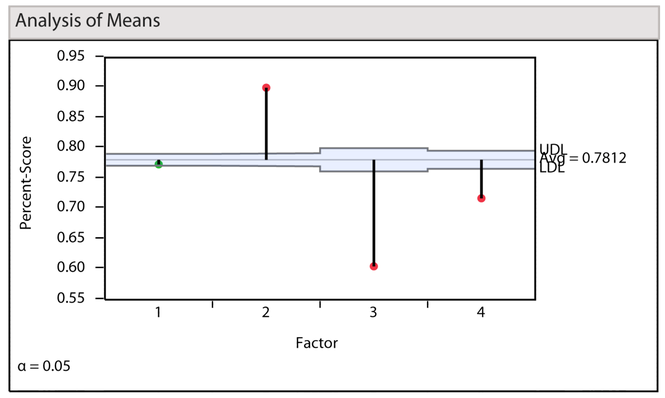


Figure 4. An analysis of means plot ranks the difficulty of factors.

DIBELS Math Pilot Study Third Grade Exploratory Factor Analysis Results

Skill set	Rotated Factor Loadings by Factor								Communality Estimates	
	1		2		3		4			
	Form A	Form B	A	B	A	B	A	B	A	B
1	-.02	-.11	**.80**	**.81**	.01	-.07	-.01	.14	**.62**	**.62**
2	.07	.18	**.65**	**.70**	.14	-.07	-.03	.00	**.54**	**.58**
3	**.84**	**.77**	-.08	.05	-.04	-.02	.08	.06	**.69**	**.65**
4	**.84**	**.79**	-.05	.04	.04	-.13	.02	.22	**.70**	**.73**
5	**.65**	**.82**	.11	.04	.16	.03	-.18	-.05	**.52**	**.68**
6	.10	.14	.14	.11	.35	.19	**.47**	**.54**	**.61**	**.54**
7	.05	.09	.06	.04	.08	-.03	**.76**	**.83**	**.69**	**.74**
8	.05	-.04	.19	.07	**.64**	*.27*	-.21	.20	.48	.16
9	.06	.02	-.11	-.10	**.84**	**.80**	.18	.17	**.81**	**.73**
Factor Variance	.29		.21		.19		.19		Model Variance = .72	

Note. N = 500. Scores were adjusted for items attempted at the student level. Factor loadings in bold represent large weights. Communality estimates in bold represent high reliability. Factor variance is the amount of variance explained in the model by the factor. Models variance represents the total variance explained by the model. Factor reliability estimates (Cronbach's Alpha) ranged from .58 to .66.

Results

New Categories and Skill Sets

Names were given to each factor related to the skill sets and ordered by difficulty:

- Factor 2 - *Moderate Multi-Number Addition Proficiency*
  - Skill Set 1 (add two two- or three-digit numbers with renaming from ones to tens and tens to hundreds).
  - Skill Set 2 (add three two- or three-digit numbers with renaming from ones to tens and from tens to hundreds).
- Factor 1 - *Moderate Multi-Digit Subtraction Proficiency*
  - Skill Set 3 (subtract a two- or three-digit number from a three-digit number with renaming from tens to ones and hundreds to tens)
  - Skill Set 4 (subtract a two- or three-digit number from a three-digit number with a zero in the tens column with renaming from tens to ones and hundreds to tens).
  - Skill Set 5 (subtract a three- or four-digit number from a four-digit number with or without renaming)
- Factor 4 - *Moderate Two-Digit by One-Digit Multiplication Proficiency*
  - Skill Set 6 (multiply a one-digit number by a two-digit number without renaming).
  - Skill Set 7 (multiply a one-digit number by a two-digit number with renaming).
- Factor 3 - *Advanced Two-Digit Dividend Division Proficiency*
  - Skill Set 8, form A (divide by a one-digit divisor resulting in a one-digit quotient and no remainder).
  - Skill Set 9 (divide by a one-digit divisor resulting in a one-digit quotient and remainder).

Discussion

Reliability of the Analysis

Similar skill sets from both forms successfully loaded onto the same factor, suggesting that items across forms maintain strong underlying relationships within the latent construct. The lone exception was skill set 8, form B. While items were randomly assigned, form A items contained *two-digit dividends* and form B items contained *one-digit dividends* , which were notably easier; the EFA suggested that the items on form B were a different skill set. Reliability is strong across factors indicating good model fit, and reliability for skill sets is moderate to strong suggesting the skill sets are well-aligned within the model.

Next Steps

Items within skill sets were categorized into subgroups based on the alignment to the Common Core State Standards for Mathematics, and new definitions were created to generate new items. These methods were repeated for all other grades, and a hierarchy of algebraic computational skill was created. **This quantitative approach provided the foundation for developing a new mathematics assessment in early childhood education: DIBELS Math.**

References

Child, D. (1990). *The essentials of factor analysis* (2nd ed.). London: Cassel Educational Limited.
Jolliffe, I.T., (2002). *Principal Component Analysis* (2nd ed.). New York: Springer-Verlag New York, Inc.
Kim, J. O., & Mueller, C. W. (1978). *Factor analysis: Statistical methods and practical issues.* Newbury Park, CA: Sage.
Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
Stevens, J. (1986). *Applied multivariate statistics for the social sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Acknowledgement

Thanks to DMG for handing me an interesting data set and asking, "Can you figure this out?" Special thanks to Rachael Latimer who listened to me for four months as I worked on it out loud.