Measures of Effective Teaching Project: The Validation Engine

Denis Newman
Empirical Education Inc.

May 2011

Empirical Education

- Research company specializing in
  - K-12 schools
  - Quantitative evaluations
  - Software tools for school systems
- Located in Palo Alto, CA

VALIDATIONENGINE

- A tool for school system administrators planning to deploy classroom observations
- Takes advantage of MET data collection and analysis
- Gives school systems web-based access to a library of videos to use in:
  - improving observation protocols
  - training of observers
- Will provide a preview of the tool *still in development*

Overview of the Process

1. Agency develops/adapts a protocol for teacher observations to use as part of evaluation.
2. Agency trains a group of practitioners (e.g., teachers, principals, specialists) to use the protocol (“raters”).
3. The Validation Engine provides an appropriate sample of videos for scoring by the raters using the protocol.
4. The raters input the scores for each of the videos.
5. The Validation Engine generates a report.

First, let’s see the Validation Engine in action

VALIDATIONENGINE

Empirical Education

Sign In

Username:

Password:

Sign In

Validation Engine Report: Three Parts

1. Inter-Rater Reliability (IRR):
   - Useful to guide rater training
2. Analysis of Protocol Composition:
   - Are some domains essentially measuring the same thing?
3. Correlation between Protocol and Benchmark:
   - Which domains are correlated to the benchmark VAM score?

FRANKLIN SCHOOL DISTRICT
CLASS ALT STUDY 1
Observation Protocol Report
April 24, 2011

Study: CLASS ALT Study 1
Description: V1.0 of Franklin School District's observation protocol
Protocol: CLASS ALT

Date study started: 3/9/2011
Number of raters: 20
Number of video assignments: 157
Number of videos rated: 30

TABLE OF CONTENTS

An example protocol

Domain	Domain Description	Range Low	Range High
ERR	Creating a learning environment of respect and rapport	1	4
ECL	Establishing a culture for learning	1	4
EM	Managing classroom procedures	1	4
ESB	Managing student behavior	1	4
EOS	Organizing physical space	1	4
CWS	Communicating with students	1	4
CQD	Using questioning and discussion technique	1	4
CEL	Engaging students in learning	1	4
CAI	Using assessment to inform instruction	1	4
CFR	Demonstrating flexibility and responsiveness	1	4
CEQ	Implementing lessons equitability	1	4

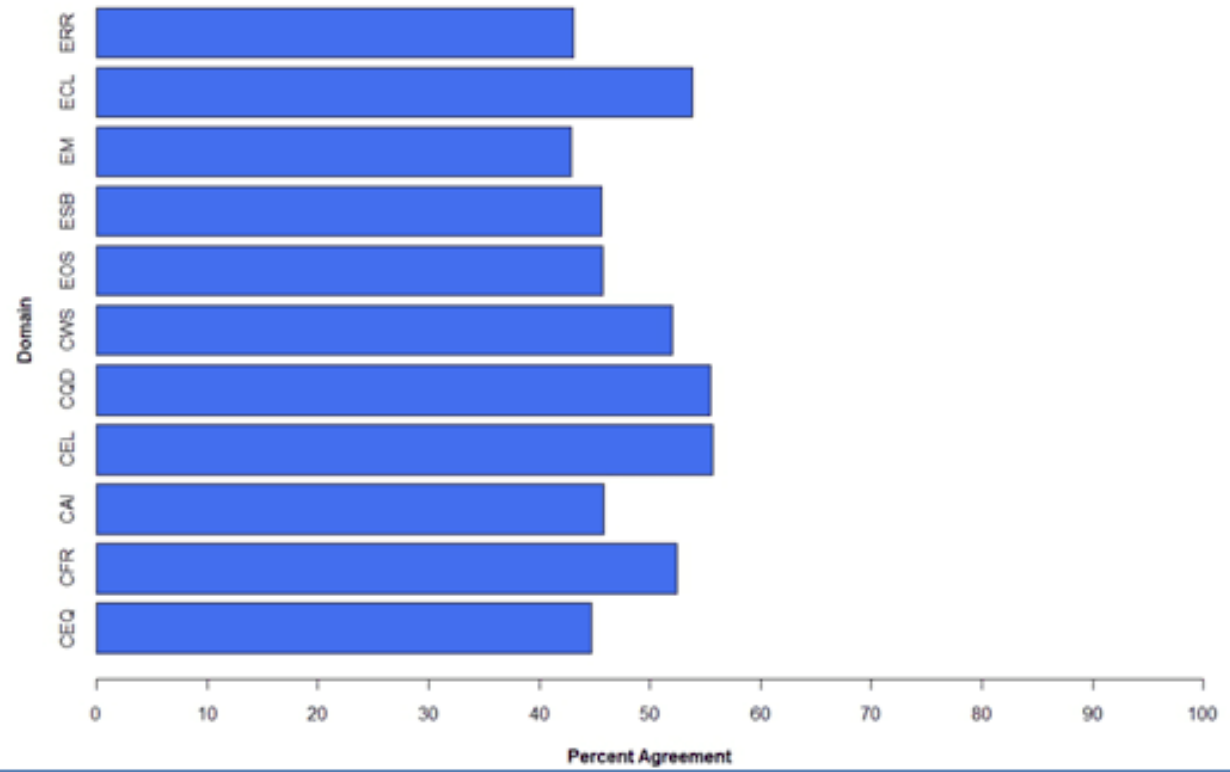
INTER-RATER RELIABILITY

Inter-rater reliability (IRR) is a measure of how consistently or similarly raters scored the videos. A high level of agreement is an important indication that the raters have been sufficiently trained and that the protocol can be reliably scored. A low level of agreement, especially if on just a few domains, may mean that parts of the protocol are ambiguous or poorly defined.

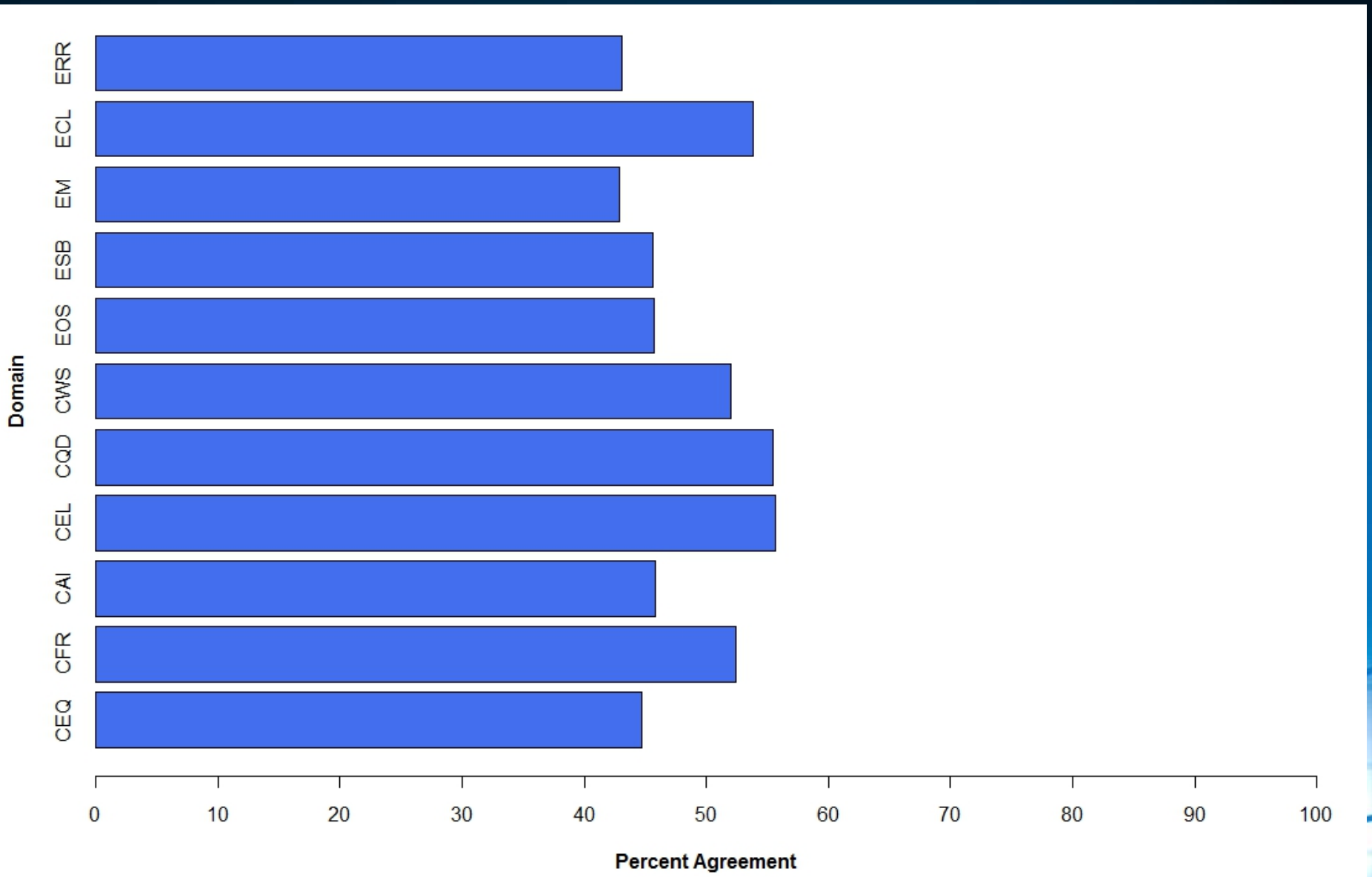
Measures in this section are calculated using a special subset of videos that every rater was asked to watch. The tool compares how raters score these videos in order to detect differences between raters and provide a measure of how consistent raters were with each other overall.

Rater Agreement by Domain

The following graph displays rater agreement by domain.



Rater Agreement by Domain



Cross Correlation of Domain Ratings

The following table describes the relationship between the ratings on different domains and may help managers determine whether the different domains in the protocol are actually measuring distinct aspects of classroom teaching or whether there is overlap or redundancy among certain domains. If two domains have a strong, positive correlation, they may actually be measuring the same aspect of teaching, just with different language. This may be an indication of redundancy in the protocol allowing for simplification by, for example, combining domains.

The value presented in each cell of the table describes the relationship between the two domains.

- A value of "+" indicates a positive relationship between two domains, i.e. teachers who are rated higher on domain A are also consistently rated higher on domain B.
- A "—" value indicates that the two domains rate teachers in a contradictory manner. Depending on the domains, this may be expected (for example, "cooperative learning" might be expected to be negatively correlated with "classroom orderliness") or it may suggest a problem with the rubric's language.
- A blank cell indicates that there is no relationship between the ratings on the two domains.

	ERR	ECL	EM	ESB	EOS	CWS	CQD	CEL	CAI	CFR	CEQ
ERR	**	+		+		+				+	+
ECL	+	**	+			+	+	+		+	+
EM		+	**	+							
ESB	+		+	**							
EOS					**						
CWS	+	+				**	+	+		+	+
CQD		+				+	**	+	+	+	+

Cross Correlation of Domain Ratings

	ERR	ECL	EM	ESB	EOS	CWS	CQD	CEL	CAI	CFR	CEQ
ERR	**	+		+		+				+	+
ECL	+	**	+			+	+	+		+	+
EM		+	**	+							
ESB	+		+	**							
EOS					**						
CWS	+	+				**	+	+		+	+
CQD		+				+	**	+	+	+	+
CEL		+				+	+	**		+	+
CAI							+		**		+
CFR	+	+				+	+	+		**	+
CEQ	+	+				+	+	+	+	+	**

Cross Correlation of Domain Ratings

	ERR	ECL	EM	ESB	EOS	CWS	CQD	CEL	CAI	CFR	CEQ
ERR	**	+		+		+				+	+
ECL	+	**	+			+	+	+		+	+
EM		+	**	+							
ESB	+		+	**							
EOS					**						
CWS	+	+				**	+	+		+	+
CQD		+				+	**	+	+	+	+
CEL		+				+	+	**		+	+
CAI							+		**		+
CFR	+	+				+	+	+		**	+
CEQ	+	+				+	+	+	+	+	**

Cross Correlation of Domain Ratings

	ERR	ECL	EM	ESB	EOS	CWS	CQD	CEL	CAI	CFR	CEQ
ERR	**	+		+		+				+	+
ECL	+	**	+			+	+	+		+	+
EM		+	**	+							
ESB	+		+	**							
EOS					**						
CWS	+	+				**	+	+		+	+
CQD		+				+	**	+	+	+	+
CEL		+				+	+	**		+	+
CAI							+		**		+
CFR	+	+				+	+	+		**	+
CEQ	+	+				+	+	+	+	+	**

Correlation between Domains and Benchmark

The following table provides descriptors about the relationship between the domain scores and the benchmark scores.

- "+" indicates a statistically significant positive relationship between a domain and the benchmark (i.e. teachers who are rated higher on this domain are also consistently rated higher on the benchmark).
- "—" indicates that the domain's rubric associates high domain ratings with lower benchmark ratings. The negative correlation is statistically significant.
- A blank cell indicates that a statistically significant relationship between the domain ratings and the benchmark was not found.

Domain	Correlation with Benchmark
ERR	
ECL	+
EM	+
ESB	
EOS	+
CWS	+
CQD	+
CEL	+
CAI	+
CFR	+
CEQ	+

Correlation between Domains and Benchmark

Domain	Correlation with Benchmark
ERR	
ECL	+
EM	+
ESB	
EOS	+
CWS	+
CQD	+
CEL	+
CAI	+
CFR	+
CEQ	+

Continuing Development

- Current functionality in beta test
- Potential future functionality
  - Protocol improvement
    - Highlighting areas that are hard to code
  - Observer training
    - Going beyond IRR to coding against an “expert”
    - Review of error patterns
- Release of Validation Engine not yet scheduled

Denis Newman
President
Empirical Education Inc.
dn@empiricaleducation.com
(650) 328-1734 x112

www.empiricaleducation.com