

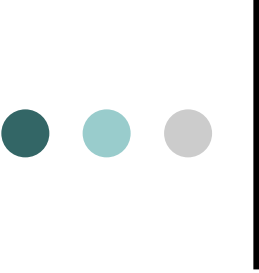


Consistency of Content Expert Q-Matrix Development

John Willse

University of North Carolina at Greensboro

Presentation at the Applying Cognitive Diagnostic Modeling in
Schools Mini-Workshop, April 24, 2008



The Q-matrix

- A Q-matrix places constraints on the model and defines the *skills/attributes* needed to answer a given question
- How is the Q-matrix constructed?
 - Confirmatory rather than exploratory
 - Relies on professional judgment or past research



Choosing the Attributes

- Possible attributes were defined for us
 - NC has a standard course of study
 - Teachers follow a pacing guide related to that course of study
- We chose to use the attributes as defined by the Department of Public Instruction
 - Students would ultimately be evaluated in relation to the course of study
 - Teachers were already familiar with those definitions and the implied skills



The Assessment

- A “benchmark” assessment was created
 - The assessment was designed to measure objectives covered up to the point of administration
 - As defined by the county pacing guide
- Through one stage of pilot testing, the assessment was refined using traditional CTT techniques
- Before using this assessment with a CDM we wanted to validate the Q-matrix
 - Verify which skills (objectives) were related to which items



Our Process

- 5 teachers familiar with the content area were recruited
- Teachers were independently asked to specify the Q-matrix
 - We didn't use *those* words, example directions to follow
- Their responses were taken as “votes”
- Any time an objective was associated with an item by at least 2 votes we chose to indicate the relationship in the Q-matrix
 - We wanted to err on the side of inclusion
 - We assumed that objectives our experts were less certain were needed could later be adjusted to be less influential during the standard setting phase



Example Directions

Part I: In general, our goal is to use a benchmark test to provide diagnostic information as to which of a set of Algebra II objectives has been mastered by each of your students. A benchmark test containing 25 items was created to measure a set of these objectives. As a first step we would like to be sure that the items in the test measure the objectives that they are intended to measure. You are going to be given a sheet of paper containing the 25 items and we ask that you provide which of the objectives that you think each item measures.



Example Directions

To better describe this procedure we use a simple example based on a test that is intended to measure basic math. Let's assume that the objectives in this class were to teach students how to add, subtract, multiply, and divide (in your case these would actually be a set of objectives that are set by the state of North Carolina). In our example, we develop a test only containing the 3 items below:



The Simple Math Example

Example Test Measuring Basic Math:

$2+3-1=?$

$2/3=?$

$2*4=?$

Notice that in this example every item does not require the four skills (add, subtract, multiply, and divide) and so we need to describe which skills are needed to answer each item. The way that we will summarize this information by using a table like the one below.

	Add	Subtract	Multiply	Divide
$2+3-1=?$				
$2/3=?$				
$2*4=?$				

We ask that you simply provide a check (or an “X”) under those skills that would be needed to correctly answer each of the items. Again we provide an example of the final table.

	Add	Subtract	Multiply	Divide
$2+3-1=?$	X	X		
$2/3=?$				X
$2*4=?$			X	



Objectives Retained for our Q-Matrix

Objectives Defining the Test:

- 1.03 Operate with algebraic expressions (polynomial, rational, complex fractions) to solve problems
- 2.01 Use the composition and inverse of functions to model and solve problems: justify results
- 2.02 Use quadratic functions and inequalities to model and solve problems; justify results
 - a. Solve using tables, graphs and algebraic properties
 - b. Interpret the constants and coefficients in the context of the problem



Objectives Retained for our Q-Matrix

Objectives Defining the Test:

- 2.04 Create and use best-fit mathematical models of linear, exponential, and quadratic functions to solve problems involving sets of data
 - a. Interpret the constants, coefficients, and bases in the context of the data.
 - b. Check the model for goodness-of-fit and use the model, where appropriate, to draw conclusions or make predictions
- 2.08 Use equations and inequalities with absolute value to model and solve problems: justify results.
 - a. Solve using tables, graphs and algebraic properties.
 - b. Interpret the constants and coefficients in the context of the problem.

The Final Q-Matrix

- The average q-matrix complexity is 1.36
 - 9 items require 2 attributes
 - 16 items require 1 attribute
- 2 of the 5 raters indicated that each item was influenced by only 1 objective

○ Stem for item 2

If one factor of $f(x) = 12x^2 - 14x - 6$ is $(2x - 3)$ what is the other factor of $f(x)$ if the polynomial is factored completely.

1.03 Operate with algebraic expressions (polynomial, rational, complex fractions) to solve problems

2.02 Use quadratic functions and inequalities to model and solve problems; justify results

The Final Q-Matrix

Item	Objective				
	1.03	2.01	2.02	2.04	2.08
1	0	1	0	0	0
2	1	0	1	0	0
3	1	0	1	0	0
4	1	0	1	0	0
5	1	0	1	0	0
6	0	1	0	1	0
7	0	1	0	0	0
8	0	0	0	0	1
9	1	0	0	0	0
10	0	0	0	0	1
11	0	0	0	0	1
12	1	0	0	0	0
13	0	0	0	0	1
14	0	0	0	0	1
15	1	0	0	0	0
16	0	0	0	1	0
17	0	0	0	1	0
18	0	0	0	1	0
19	0	0	0	1	0
20	1	0	1	0	0
21	1	0	0	1	0
22	0	0	0	0	1
23	0	0	0	1	0
24	1	0	1	0	0
25	1	0	0	1	0



G-Theory, 45 Second Primer

- Generalizability (G) Theory addresses the dependability of measurements
 - Dependability is G-Theory's analog to classical test theory's (CTT) concept of reliability
- G-theory is most easily thought of as an extension of CTT
 - In CTT only one source of error is considered at a time (e.g., internal consistency OR alternate form)
 - In G-theory multiple sources of error can be incorporated into a single coefficient (g-coefficient or phi-coefficient)
- Dependability (D) studies are G-theory's analog to CTT's Spearman-Brown prophecy formula
 - D-studies allow us to estimate dependability under a variety of alternative measurement scenarios
 - What happens if we add more raters?



G-theory and the Q-Matrix

- The Q-matrix was specified based on our teachers' responses
 - Teacher did not perfectly agree
- We wanted to measure whether we had a dependable estimate of the Q-matrix
 - Dependability (i.e., reliability) being a prerequisite to validity
 - And we wanted to estimate how many teacher raters would be needed (fewer or more) to achieve adequate dependability



G-Study Design

Sources of Variability

- Test-Items, Object of Measurement

Sources of Error Variance

- Raters, Teachers indicating which objectives are required in order to answer items
- Objectives, Attributes Influencing Items
 - Objectives were treated as fixed
 - Meaning not all associated variance would be considered error
 - The tested objectives defined the domain (rather than being sampled from it)



Problems with Analysis

- In the g-study (estimation of variance components) several of the estimated components were negative
 - A fairly common occurrence in g-studies with small samples
 - Here 25 objects of measurement
- We had decided to treat objectives as fixed, so analyzing the dependability of each objective separately was a reasonable solution

Dependability of the Q-Matrix

- In g-theory there is a coefficient for relative decisions (i.e., ranking), g , and one for absolute decisions (i.e., criteria-based), Φ
 - We interpreted Φ
- Under our current design (shaded row) we had adequate dependability with objective 2.01 and 2.08

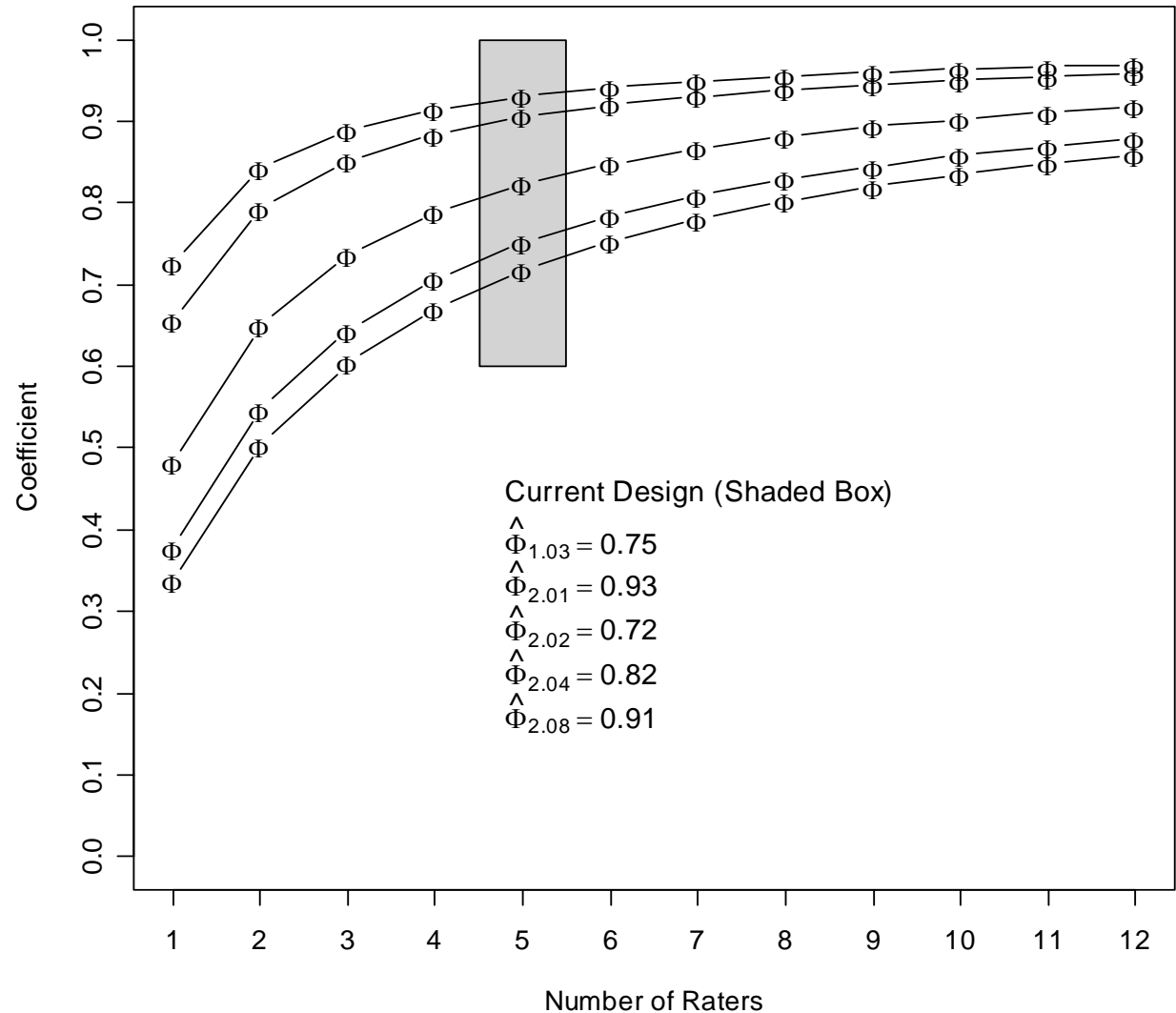
Table 1. Dependability of Assignments

Raters	Objective				
	1.03	2.01	2.02	2.04	2.08
1	0.38	0.73	0.34	0.48	0.66
2	0.55	0.84	0.50	0.65	0.79
3	0.64	0.89	0.60	0.74	0.85
4	0.71	0.91	0.67	0.79	0.88
5	0.75	0.93	0.72	0.82	0.91
6	0.78	0.94	0.75	0.85	0.92
7	0.81	0.95	0.78	0.87	0.93
8	0.83	0.96	0.80	0.88	0.94
9	0.84	0.96	0.82	0.89	0.95
10	0.86	0.96	0.84	0.90	0.95
11	0.87	0.97	0.85	0.91	0.95
12	0.88	0.97	0.86	0.92	0.96

Improving Dependability

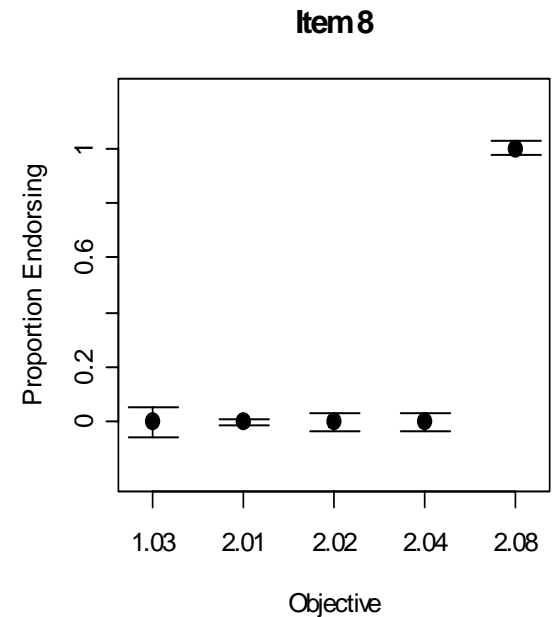
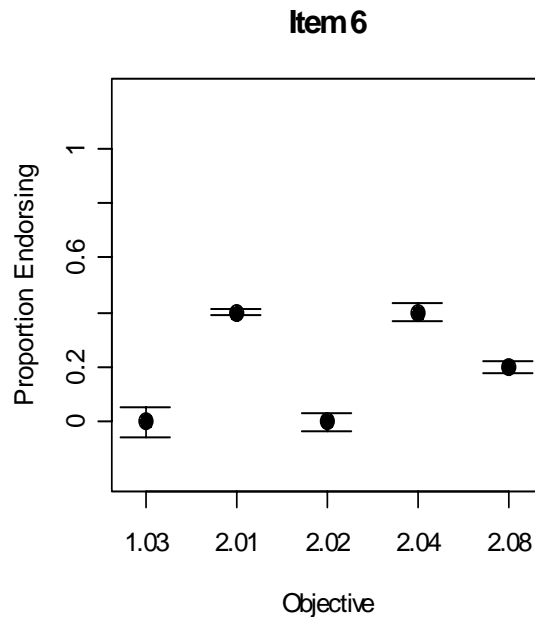
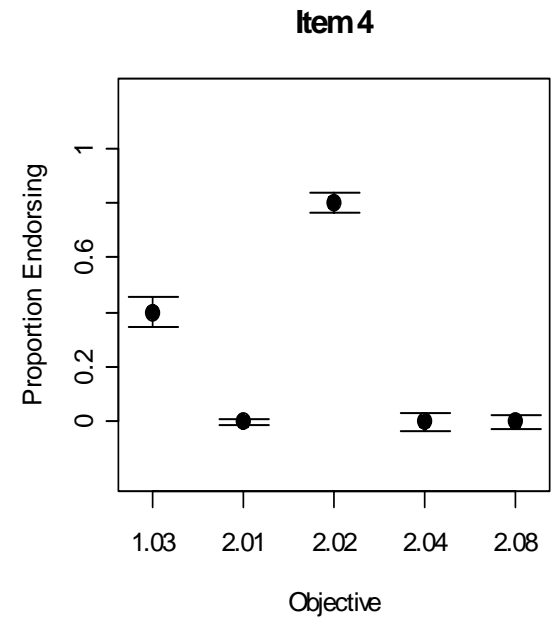
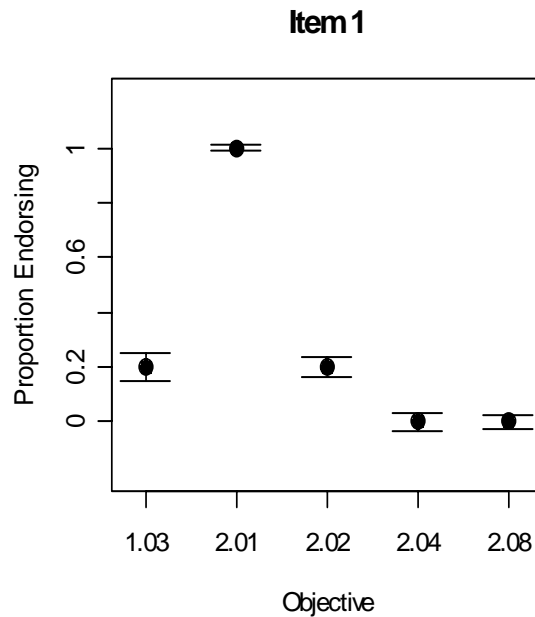
- If we recruit raters and implement the procedure in a similar manner we would need twice as many
- Clarifying our procedures might be a more efficient approach
 - More precise definition of attributes
 - More clarity in description of task

Dependability of Assignment by Objective



A Few Words on SEM

- We could also phrase the discussion of dependability using standard errors
- Using standard errors we see that even under our current measurement design, 95% confidence intervals are reasonable
- Rules for including an objective in the Q-matrix could be based on whether the lower bound of a CI was above a threshold





Lessons for Moving On

- When we perform this task in the future we will
 - Recruit more teachers to produce the q -matrix
 - Spend more time talking about the *meaning* of the objectives that are used as attributes
 - Make sure that directions regarding complex structure are clear
 - Consider revising rules regarding proportion of “votes” objectives must receive to be included in the Q-matrix