

The Vertical Scaling of Science Achievement Tests

Mark D. Reckase
Joseph Martineau
Michigan State University

Measuring Growth

- When tests are administered every year, there is a natural desire to look at change in performance.
- Determining change in performance requires tests that are vertically scaled.
- Numerous psychometricians have said that vertical scaling is impossible.

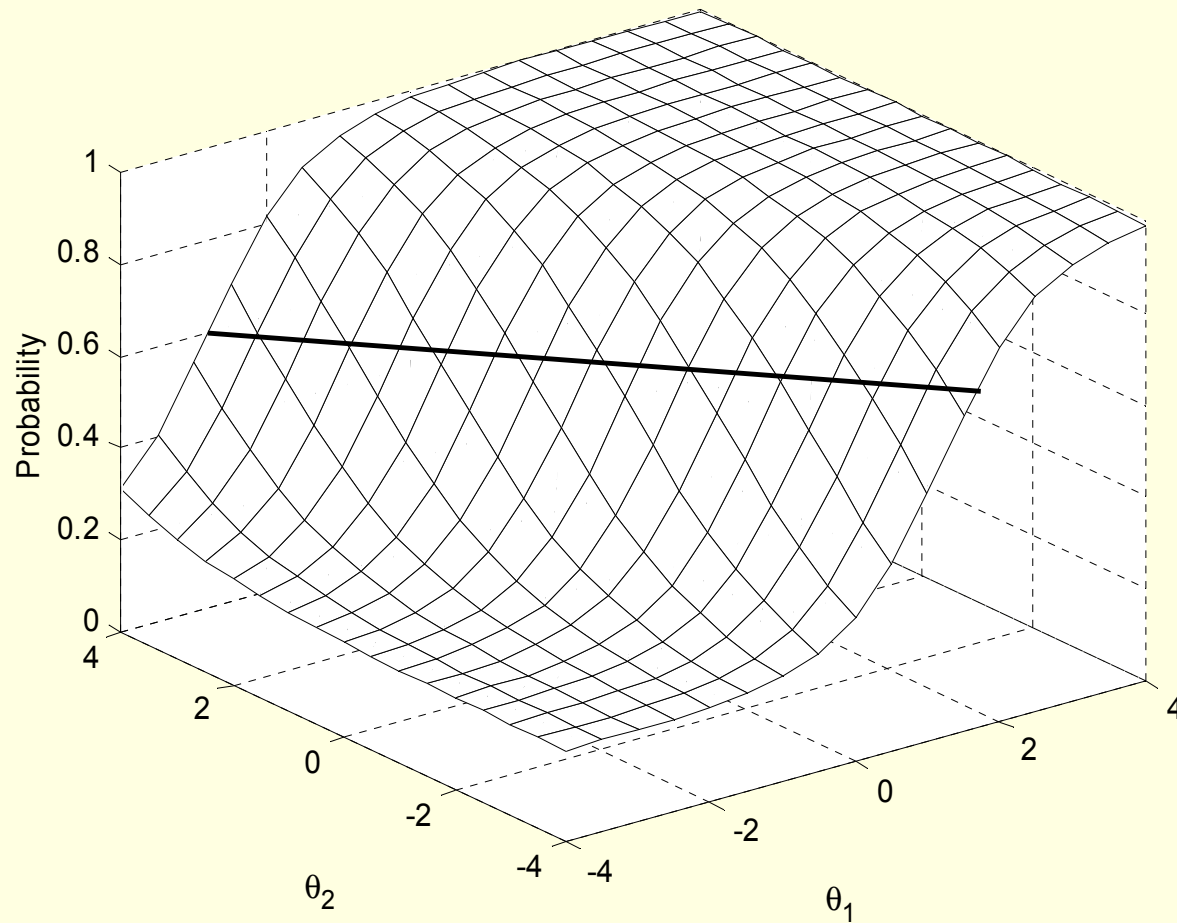
Multidimensional Modeling of Growth

- The reason that vertical scaling is impossible is that the tests are not parallel.
 - Measure different constructs
 - Have different difficulty
- Can solve the different construct problem by embedding the test in a high dimensional space.
- Difficulty is dealt with in the usual IRT way.

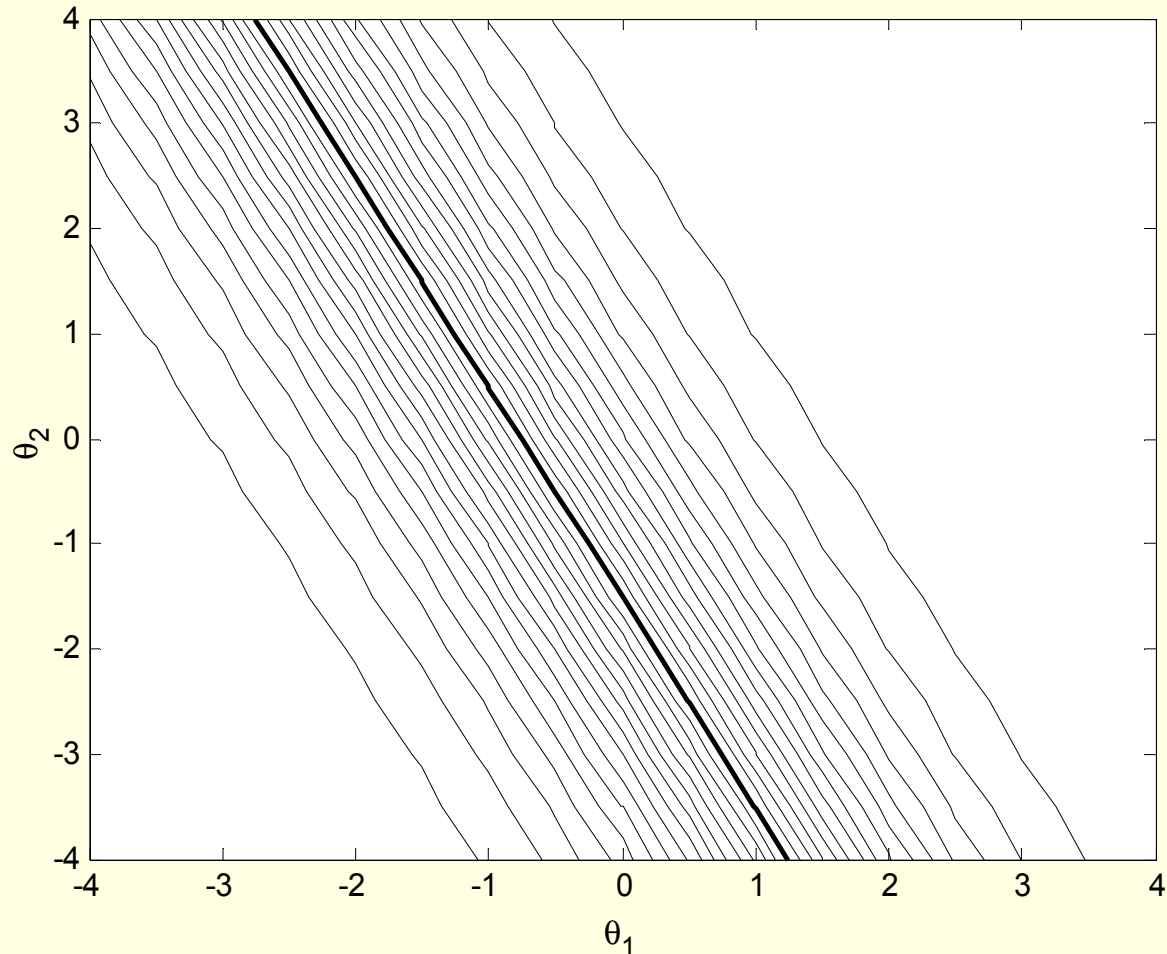
The Model

$$P(u_{ij} = 1 | \vec{\theta}_j, \vec{a}_i, d_i, c_i) = c_i + (1 - c_i) \frac{e^{\vec{a}_i' \vec{\theta}_j + d_i}}{1 + e^{\vec{a}_i' \vec{\theta}_j + d_i}}$$

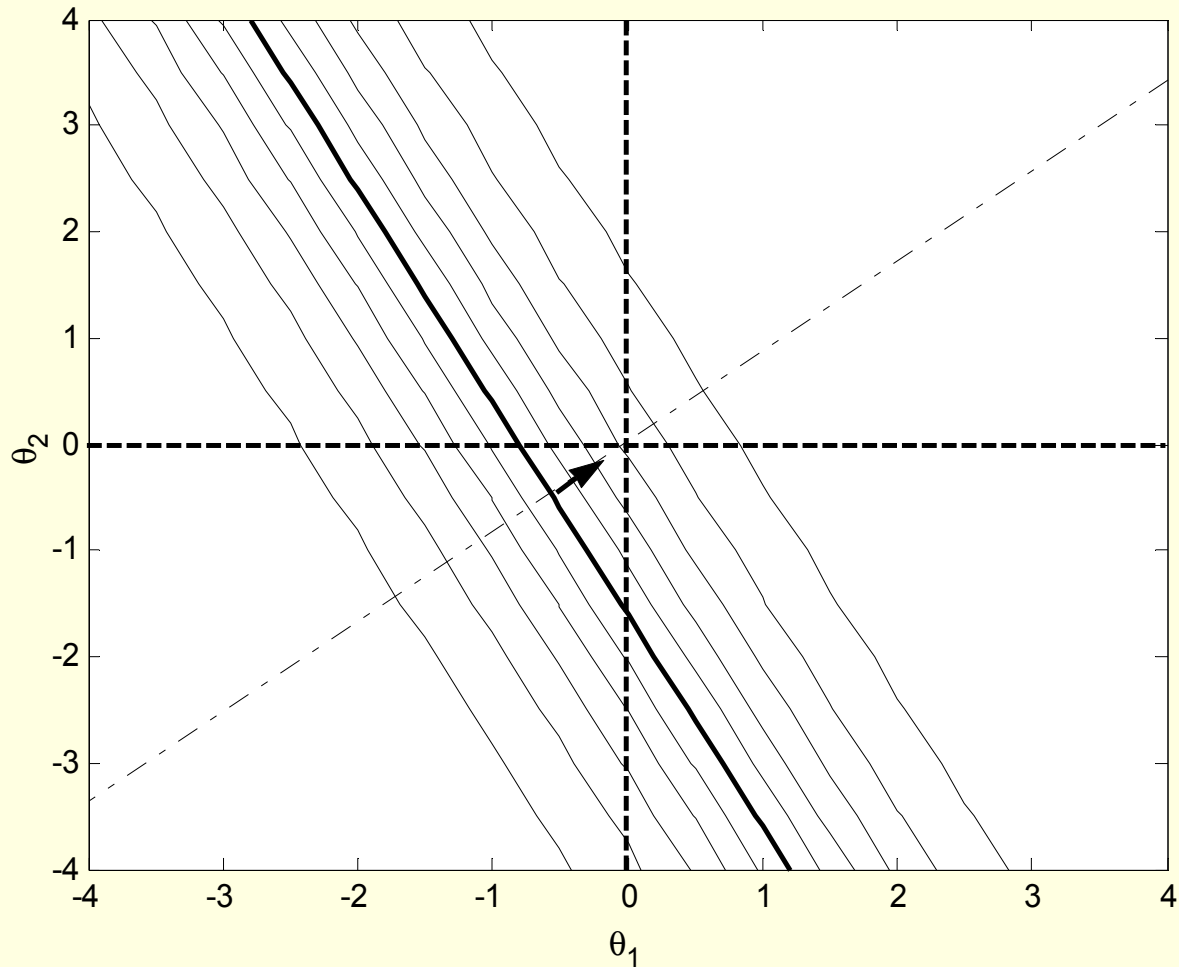
The Model Represents a Surface in a Multidimensional Space



There Are Different Ways of Representing the Surface



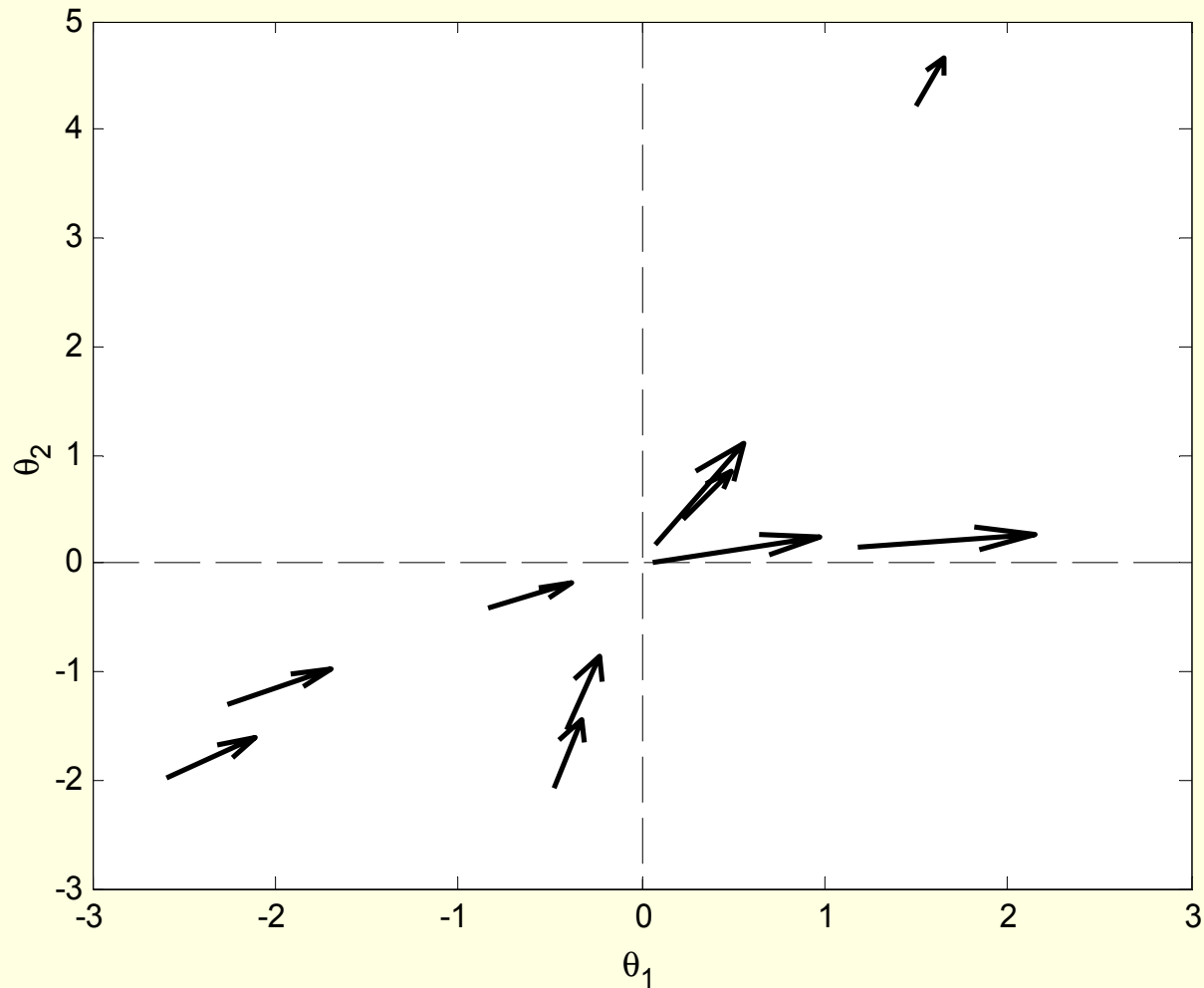
Vector Representation



Using the Vector Representation

- It is difficult to represent a lot of items using the surface or contour representation of an item.
- The vector approach allows many items to be shown on one graph.
 - Vectors that are pointing in the same direction are measuring the same combination of skills and knowledge.
 - The base of the vector indicates the difficulty of the item.

Vector Plot for Items Measuring Two Constructs



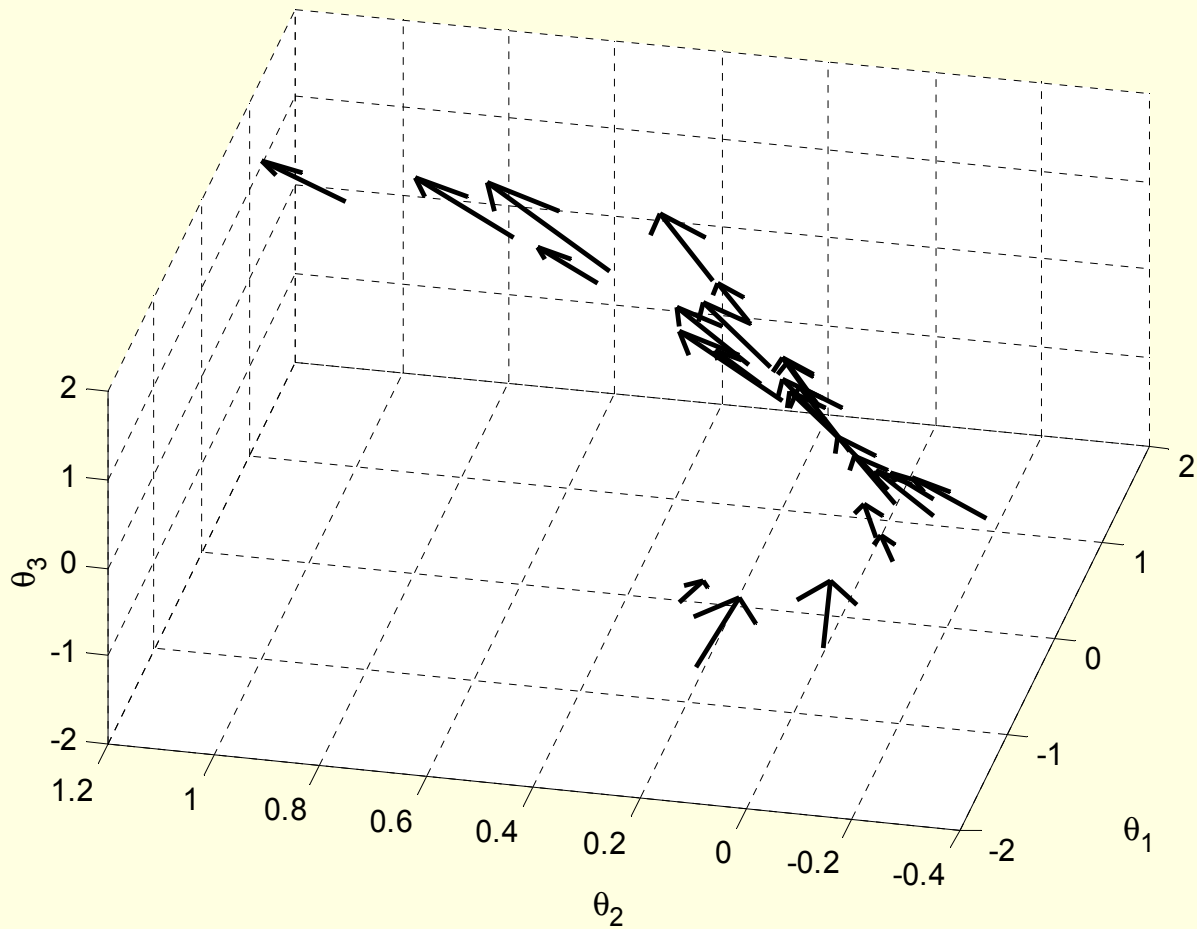
Purpose of the Analysis

- Link together grade level science tests into a common multidimensional space
- Show how the science tests at different grade levels measure different constructs
- Show how growth in science over grade levels does not occur uniformly over all of the skills and knowledge.

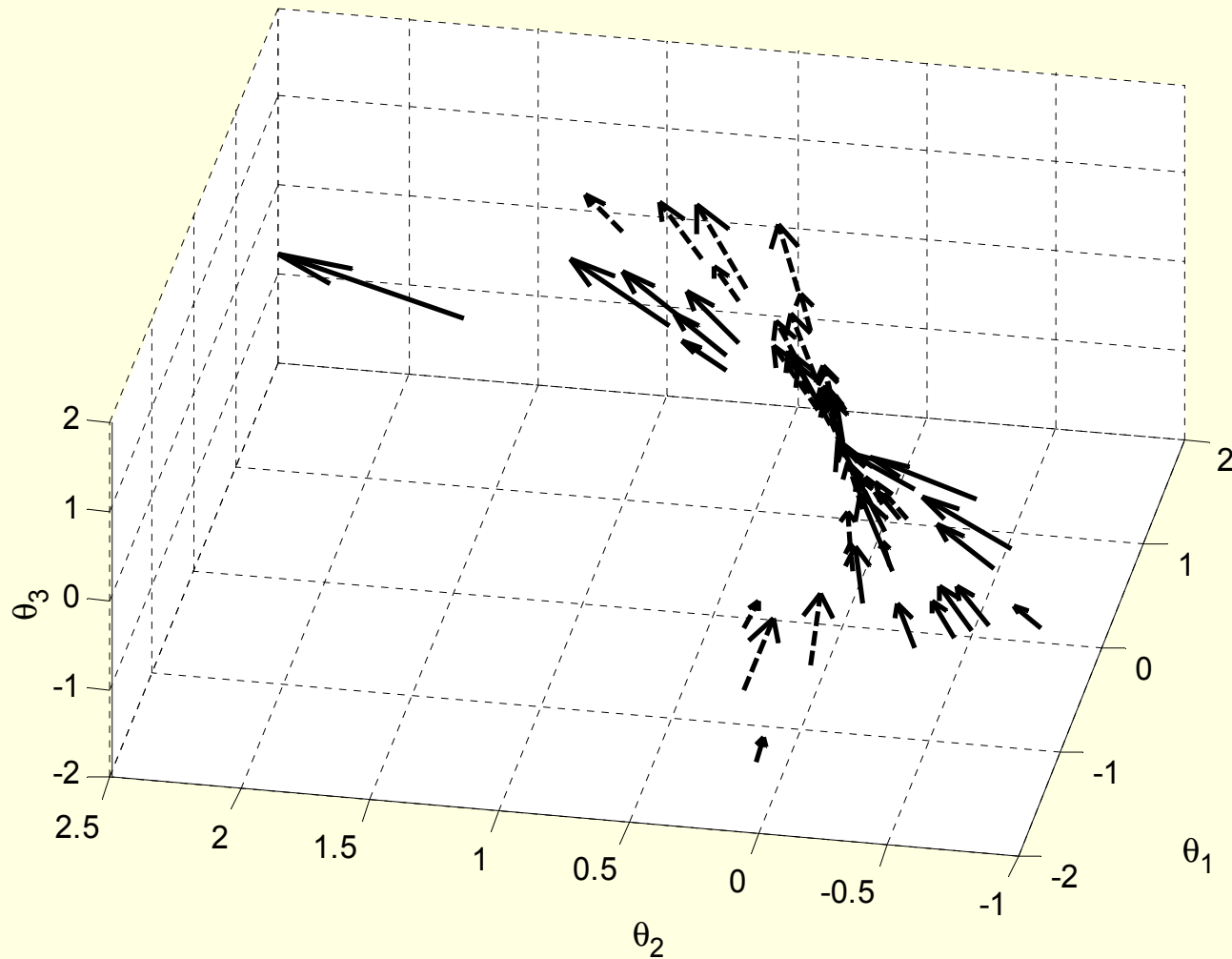
The Data

	Level 16	Level 17	Level 18
Grade 5	2408		
Grade 6	2044	2044	
Grade 7		1734	1734
Grade 8			1909

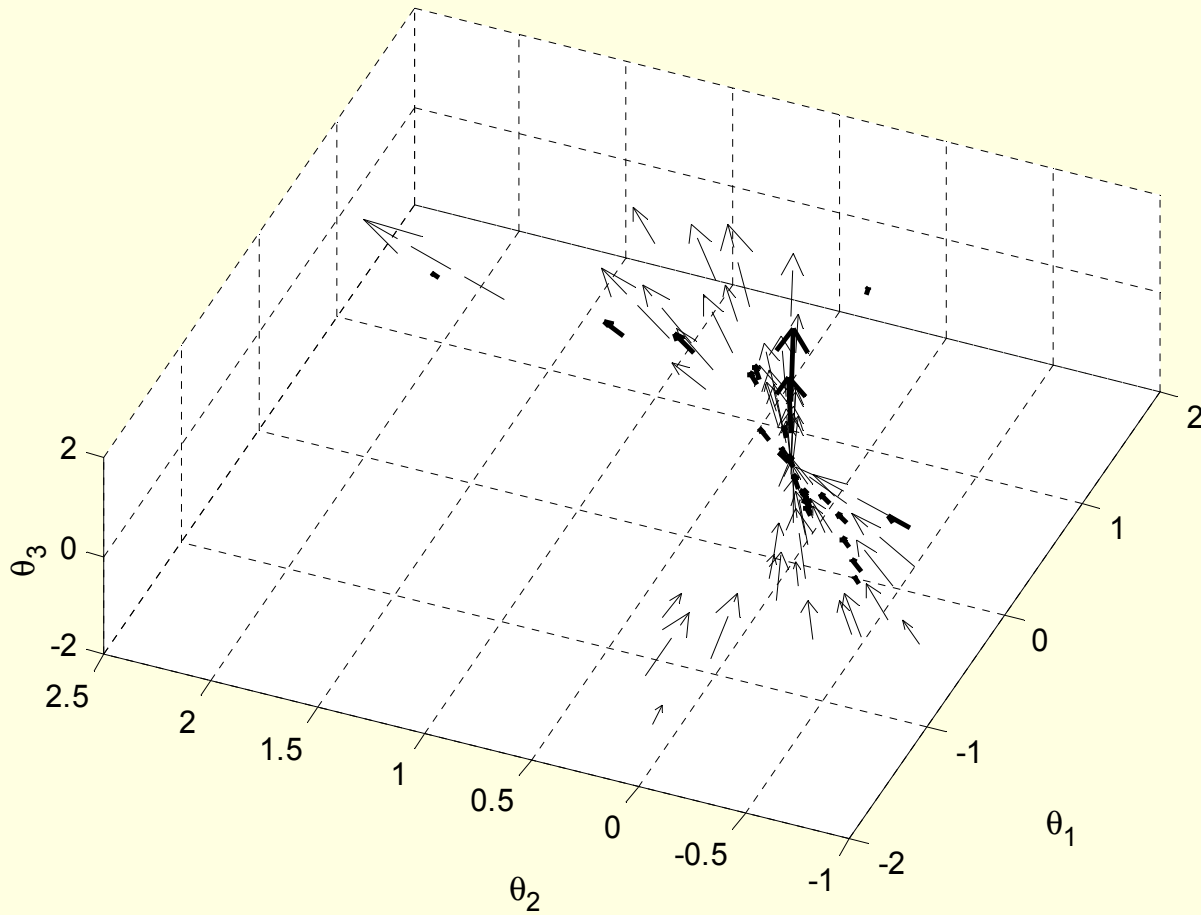
Items for Level 16



Items for Levels 16 and 17



Items for Levels 16, 17, and 18



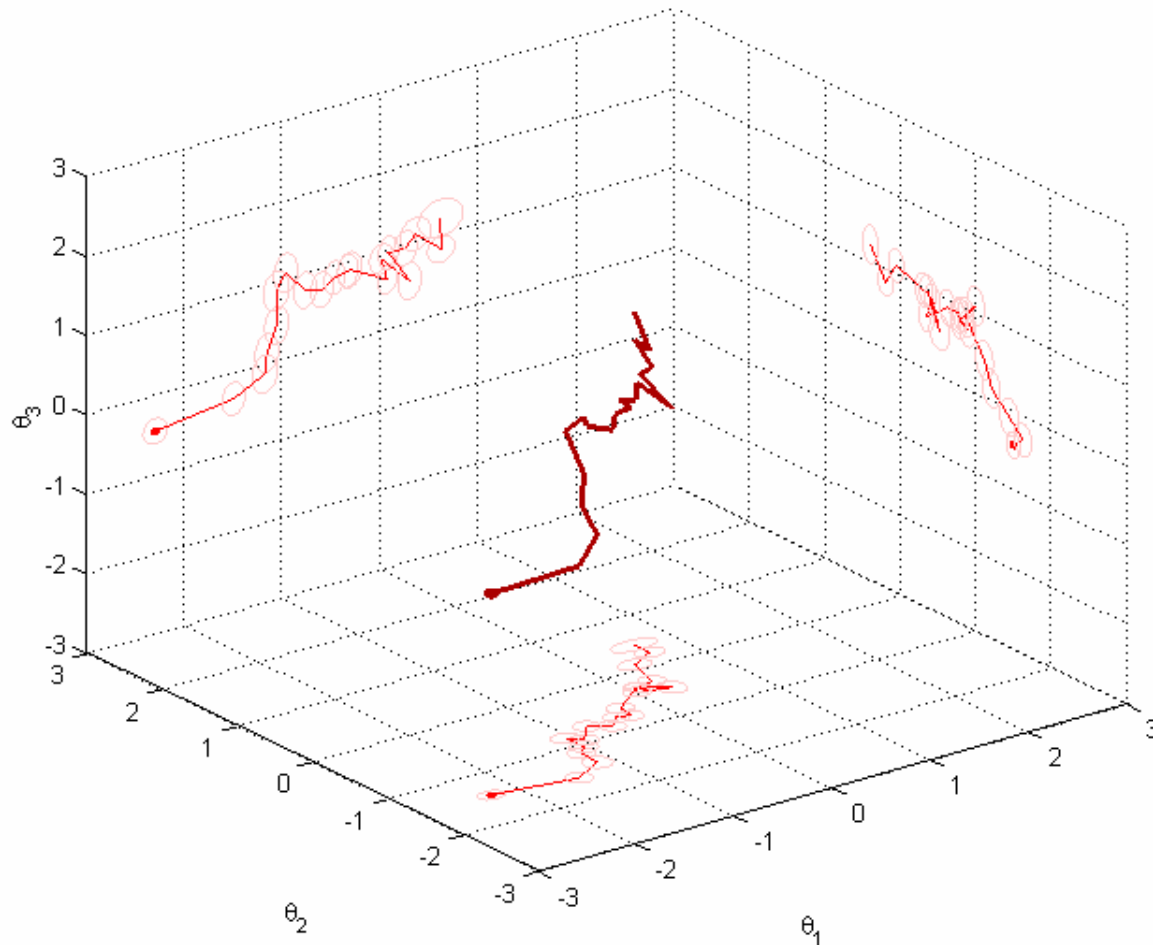
Item Plots

- Difficult items are sensitive to different combinations of constructs than the easy items.
- Patterns of sensitivity to the constructs are not the same for the different levels of tests.
- These plots are an oversimplification of reality. More dimensions are needed to represent the item characteristics.

Change in Student Performance

- To show the change in student performance, the three levels of the test were scaled using a unidimensional IRT model.
- The proficiency estimates were sorted into twenty groups. The multidimensional vectors for the students were averaged for each group.
- The results are plotted to show the change in skills and knowledge as students improve in performance.

Student Performance from Grades 6 through 8



Change in Student Performance

- There is an initial large gain along Dimension 1.
- Then there is a large increase along Dimension 3.
- Then there is an increase along a combination of Dimensions 1 and 2.
- Growth does not occur uniformly along all of the dimensions.

Next Steps

- Get test items so we can give meaning to the dimensions.
- Do a high dimensional analysis to avoid problems with projecting solutions to lower dimensional spaces.
- Produce final version.

Conclusions

- Assessing growth in science achievement is very complex because of the breadth of knowledge and skills in curriculum frameworks.
- Multidimensional IRT can be used to model this complexity.
- Test items must be sensitive to the multitude of dimensions to be able to show the nuances of growth.
- Tests should be carefully designed for multidimensional IRT analysis.