



**Dynamic Interactive Formative Assessment Tasks and End-of-Unit Tests for Measuring Challenging Concepts and Skills of Diverse Middle School Students**  
**Technical Report**

PR/Award# S368A150019

William Auty and Laura J. Wright

# Technical Report

---

08/21/2020

## 1. Table of Contents

1. TABLE OF CONTENTS.....	1
2. ONPAR OVERVIEW.....	4
3. TASK DESCRIPTIONS .....	6
3.1. MOUNTAIN HIKE .....	6
3.2. WORKOUT .....	6
3.3. BODY SYSTEMS EOU.....	6
3.4. SCIENCE NIGHT RESCORED.....	6
3.5. CAMPING AND CHEMISTRY RESCORED.....	6
3.6. CHEMICAL REACTIONS EOU RESCORED.....	6
3.7. FOOD WEBS.....	7
3.8. ABIOTIC.....	7
3.9. ECOLOGY EOU.....	7
3.10. ENERGY IN THE WORLD .....	7
3.11. ENERGY IN ACTION.....	7
3.12. ENERGY EOU .....	7
3.13. FARM GENETICS .....	7
3.14. MARIGOLDS AND MOSQUITOES.....	8
3.15. HEREDITY EOU .....	8
3.16. LIGHT AND WAVES AT THE AQUARIUM.....	8
3.17. PICTURE DAY.....	8
3.18. LIGHT EOU.....	8
3.19. SCHOOL GARDEN .....	8
3.20. LIVING THINGS AND FOOD .....	9
3.21. MATTER AND ENERGY IN LIVING SYSTEMS EOU.....	9
3.22. SPACESTATION.....	9
3.23. APPLES .....	9
3.24. MOTION AND FORCES EOU .....	9
3.25. MATTER IN NATURAL RESOURCES RESCORED .....	9
3.26. MATTER IN THE KITCHEN RESCORED .....	9
3.27. PARTICLE NATURE OF MATTER EOU RESCORED.....	10
3.28. CHANGING CONTINENTS .....	10
3.29. DYNAMIC EARTH .....	10
3.30. PLATE TECTONICS EOU.....	10
3.31. WATER ON THE MOVE.....	10
3.32. WHERE'S WATER? .....	11
3.33. WATER AND LANDFORMS EOU .....	11
3.34. WEATHER AT THE BEACH .....	11

3.35. STORMS.....	11
3.36. WEATHER EOU.....	11
4. DESCRIPTIVE STATISTICS FOR ALL TASKS.....	12
4.1. NUMBER OF STUDENTS ASSESSED .....	12
4.2. TOTAL SCORE FREQUENCIES .....	14
4.3. TASK RELIABILITIES (CRONBACH'S ALPHA AND IRT) .....	15
5. RELIABILITY AND ITEM STATISTICS FOR EACH TASK .....	16
5.1. MOUNTAIN HIKE.....	16
5.2. WORKOUT .....	19
5.3. BODY SYSTEMS EOU .....	21
5.4. SCIENCE NIGHT RESCORED.....	24
5.5. CAMPING AND CHEMISTRY RESCORED.....	26
5.6. CHEMICAL REACTIONS EOU RESCORED .....	28
5.7. FOOD WEBS.....	31
5.8. ABIOTIC .....	34
5.9. ECOLOGY EOU.....	37
5.10. ENERGY IN THE WORLD .....	40
5.11. ENERGY IN ACTION.....	43
5.12. ENERGY EOU.....	45
5.13. FARM GENETICS .....	48
5.14. MARIGOLDS AND MOSQUITOES .....	50
5.15. HEREDITY EOU.....	53
5.16. LIGHT AND WAVES AT THE AQUARIUM .....	56
5.17. PICTURE DAY .....	59
5.18. LIGHT EOU.....	61
5.19. SCHOOL GARDEN.....	63
5.20. LIVING THINGS AND FOOD .....	65
5.21. MATTER AND ENERGY IN LIVING SYSTEMS EOU .....	67
5.22. SPACESTATION.....	69
5.23. APPLES.....	72
5.24. MOTION AND FORCES EOU.....	75
5.25. MATTER IN NATURAL RESOURCES RESCORED .....	78
5.26. MATTER IN THE KITCHEN RESCORED .....	81
5.27. PARTICLE NATURE OF MATTER EOU RESCORED .....	84
5.28. CHANGING CONTINENTS.....	86
5.29. DYNAMIC EARTH.....	89
5.30. PLATE TECTONICS EOU .....	92
5.31. WATER ON THE MOVE .....	95
5.32. WHERE'S WATER?.....	98
5.33. WATER AND LANDFORMS EOU .....	101
5.34. WEATHER AT THE BEACH .....	104
5.35. STORMS .....	107
5.36. WEATHER EOU.....	110

**6. VALIDITY STUDIES ..... 113**

<b>6.1. EXTERNAL VARIABLES.....</b>	<b>113</b>
6.1.1. MODEL THE EFFECTS OF ALL EXTERNAL FACTORS ON ONPAR TASK RESULTS.....	114
6.1.2. APPLY THE MODEL TO ALL TASKS. ....	116
6.1.3. WHAT IS THE CONTRIBUTION OF THE MODEL EFFECTS FOR EACH TASK?.....	118
6.1.4. SCIENCE MODELS.....	121

## **2. Project Overview**

This report provides technical information on the *Dynamic Interactive Formative Assessment Tasks and End-of-Unit Tests for Measuring Challenging Concepts and Skills of Diverse Middle School Students*, a federally funded enhanced assessment project with a particular focus on science for all. Over a five-year period, the project has researched and developed a system of technology enhanced science assessments for middle school students aligned with the Next Generation Science Standards (NGSS); a suite of 12 units of middle school science materials were developed, piloted, and psychometrically evaluated.

The project uses an assessment methodology called ONPAR which relies upon multisemiotic representations to both present assessment items and open up response types. The theoretical underpinnings and empirical support for the assessment methodology come from the fields of semiotics, cognitive science, gaming, and Evidence Centered Design (e.g., Gee, 2004, 2007; Graf & Kinshuck, 2008; Jewitt, 2008; Kane, 2013; Kress, 2003; Kress, 2010; Kress & van Leeuwen, 2001; Lohr & Gall, 2008; Mislevy, 2009; Mislevy, 2013).

To create the tasks, assessment developers started with an examination of the NGSS to identify valued knowledge, skills. In addition, task developers also used Learning Progression research to identify appropriate grade-level expectations for each of the content units (Hadenfelt, Liu, & Neumann, 2014; Maloney, Confrey, & Nguyen, 2014; Smith, Wiser, Anderson, & Krajcik, 2006; Steedle & Shavelson, 2009). Once measurement goals and valued evidence were identified, multisemiotic tasks were drafted and scoring schemes were created and formalized. ONPAR assessments are 8-14 screens long and provide an opportunity to assess student understanding of all three dimensions of the NGSS, disciplinary core ideas, science and engineering practices, and crosscutting concepts. All tasks were professionally designed and programmed in HTML 5 for administration on a standard web browser. Tasks were piloted across the United States from 2017-2019.

While tasks in the project were automatically scored for both numeric scoring and diagnostic evaluations, this report provides information only on the numeric scoring outcomes from the assessments. Information on the diagnostic reporting of the tasks is provided in an additional report (see Winter, 2020 for a detailed evaluation). As part of the project, all tasks were revised based on pilot outcomes. After revisions were made, data from a sample of six tasks were rerun with revised scoring in place to determine whether revisions would be successful in improving tasks. These tasks are labeled in the report as ‘rescored.’

In the section that follows, written descriptions of the classroom-embedded tasks and end-of-unit tests are provided. Information includes the contextual setting of the assessments as well as measurement goals. Section 4 contains descriptive statistics for all tasks including the number of students tested, score frequencies and reliability statistics. Section 5 breaks down statistical information for each task

including the item response frequencies, item-total score correlations, response step difficulties for the polytomous items, and item fit. Section 6 contains information on the validity studies conducted for the project. Project outcomes were modelled using the external variables of teacher experience, English Learner (EL) status, state math, ELA, and science scores, ethnicity, and socio-economic status.

### **3. Task Descriptions**

#### **3.1. MOUNTAIN HIKE**

Within the context of a mountain hike, students demonstrate understanding of how body structures break down and transport food and what happens when food gets to a cell. Next, students analyze and interpret data on breathing and heart rate during different activities, construct graphs to show relationships and make a claim about these relationships. Finally, students explain how altitude affects body systems.

#### **3.2. WORKOUT**

Students demonstrate their understanding of systems, structures and functions, and the levels of organization in our bodies. They relate structure and function within a system and also between systems, especially as it relates to exercising and eating food.

#### **3.3. BODY SYSTEMS EOU**

Students demonstrate their understanding of content assessed previously in the Mountain Hike and Workout tasks. The emphasis in this task is on structure and function in cells and body systems to explain how eating and exercise are related. In addition, students demonstrate their understanding of constructing explanations and developing and using models.

#### **3.4. SCIENCE NIGHT RESCORED**

Students are engaged in three investigations following a Science Night theme. First, students compare the properties of starting and ending substances to demonstrated whether they can make a conclusion about whether or not a chemical reaction has taken place, explain the process of a chemical reaction, and model starting and ending substances in an investigation.

#### **3.5. CAMPING AND CHEMISTRY RESCORED**

Students demonstrate their understanding of the process of chemical reactions within the context of camping. First, students demonstrate understanding of conservation of matter within a chemical reaction. Then they revise a model and generate their own model of the chemical reaction process. Finally, students construct explanations about the steps within reactions.

#### **3.6. CHEMICAL REACTIONS EOU RESCORED**

Students are engaged in explorations of how two different metals chemically react with other substances. First, students demonstrate understanding of how to model parts of a chemical reaction and conserve matter in the process. Then, students model starting and ending substances and construct an explanation about chemical reactions. Finally, students build a model of the step by step process within a chemical reaction, and they demonstrate understanding of how products form.

### **3.7. FOOD WEBS**

Students determine how shifts in biotic factors affect populations in a forest food web. Students show how the removal of a top predator, an insect infestation in trees, and the introduction of an invasive species, can affect a forest ecosystem.

### **3.8. ABIOTIC**

Students determine how biotic and abiotic changes in an aquatic environment affect organisms in that ecosystem. Students also show how human actions can impact the environment.

### **3.9. ECOLOGY EOU**

Students demonstrate understanding of interdependent relationships in two different ecosystems, an ocean ecosystem and a forest ecosystem. Students demonstrate understanding of biotic and abiotic factors and how they might affect populations in an aquatic food web. Students also show an understanding of matter and energy flow in organisms; they determine how shifts in factors affect populations in a forest ecosystem.

### **3.10. ENERGY IN THE WORLD**

Students visualize the energy present in a system and identify different types of energy; they establish the relationship between potential energy, kinetic energy, mass and distance prior to planning an investigation involving these relationships. Students also interpret and graph data from an investigation about kinetic energy, distance and mass.

### **3.11. ENERGY IN ACTION**

Students demonstrate their understanding of the types of energy within four different “athletic” scenarios: a trampoline, a pole vault, a basketball, and bowling. They demonstrate understanding of energy present in these systems and that they can track energy transfers. Students interpret graphs and represent data in graphical displays.

### **3.12. ENERGY EOU**

Students demonstrate their understanding of energy relationships in systems, analyzing and interpreting energy transfer data, and constructing explanations about how energy transforms. In addition, students demonstrate their understanding of energy at the molecular level within a chemical reaction.

### **3.13. FARM GENETICS**

Students analyze plant and animal traits on a farm and consider the offspring that result from parents with different variations of a trait. Students explain the

difference in offspring in asexual and sexual reproduction and use this to explain how genetic variation in the two types of reproduction lead to changes over time. Finally, students predict what genotypes and phenotypes are possible with two corn plants for two different traits, kernel color and shape.

### **3.14. MARIGOLDS AND MOSQUITOES**

Students demonstrate their understanding of natural selection and how genetic traits can change over time. Students construct explanations about how populations change in three different cases.

### **3.15. HEREDITY EOU**

Students demonstrate their understanding of inheritance and variation of traits, growth and development of organisms, natural selection and adaptation. In addition, students demonstrate their ability to construct explanations and develop and use models.

### **3.16. LIGHT AND WAVES AT THE AQUARIUM**

A visit to the zoo's aquarium becomes an exploration of student understanding of light behavior. Students demonstrate their ability to model pathways of light, show how light interacts with objects and surfaces, and apply their understanding of waves to demonstrate how light behaves as it moves between media.

### **3.17. PICTURE DAY**

Students explore how light interacts with and around different screen surfaces. Students demonstrate their understanding of how differently colored light affects the appearance of things, create a model of how the eye sees color and use what they know about wavelengths to explain the effect of filters on light.

### **3.18. LIGHT EOU**

Students demonstrate their understanding of light and sound waves in a museum setting. Students correlate their understanding of sound waves with pitch and volume, use models of waves to explain color and the transmission of light through materials, and develop models of how light interacts with objects and screens.

### **3.19. SCHOOL GARDEN**

Students analyze whether or not certain data should be used as evidence to explain how light affects plant growth; they explain the results of experiments about photosynthesis as well as the process of photosynthesis and how it helps plants grow.

### **3.20. LIVING THINGS AND FOOD**

Students demonstrate their understanding of the process of respiration as a process within living organisms and in a natural system, construct explanations about the inputs and outputs of respiration, and revise a model.

### **3.21. MATTER AND ENERGY IN LIVING SYSTEMS EOU**

Students demonstrate their understanding of both photosynthesis and respiration, their ability to plan investigations, construct explanations, and develop and use models.

### **3.22. SPACE STATION**

Students demonstrate understanding of forces and motion in the context of objects in space, including concepts relating to gravity. They demonstrate the ability to create explanations, plan an investigation and analyze and interpret data.

### **3.23. APPLES**

Students demonstrate understanding of motion and forces in the context of wagons moving apples. Students apply concepts relating to motion and force to graphical analysis and interpretation. Students develop system models, and describe forces and motion using various graphs.

### **3.24. MOTION AND FORCES EOU**

Students demonstrate understanding of motion and forces in the context of moving objects on Earth and forces of gravity in space. Students apply concepts relating to motion and force to graphical analysis and interpretation. They design an investigation, develop system models, and create explanations.

### **3.25. MATTER IN NATURAL RESOURCES RESCORED**

Natural resources serve as the backdrop for students to demonstrate understanding of the particle nature of substances. Students demonstrate their understanding that matter, in various forms, is what has mass and takes up space. Then, they analyze and interpret various simple molecular models of matter, including a model of a substance prior to and after a change in temperature. Students also apply their understanding of characteristic properties of matter to solve for an unknown.

### **3.26. MATTER IN THE KITCHEN RESCORED**

In the context of changes in matter in a kitchen, students describe and model particles of water as they change phases. They also identify a question, revise conditions, and make predictions for an investigation related to heating a substance. Students predict and explain how diffusion is influenced by heat.

### **3.27. PARTICLE NATURE OF MATTER EOU RESCORED**

Students demonstrate understanding of changes in matter in a party setting; they model gases in balloons to demonstrate knowledge of pure substances and mixtures, as well as how temperature affects gases. Students examine the phenomenon of wax candles burning to show how particles in a wax candle are affected during a phase change. Students also interpret an investigation about the change in temperature that occurs when candles are put out in different amounts of water and use a model to show what is happening to the particles in different conditions of the investigation.

### **3.28. CHANGING CONTINENTS**

Students analyze data from fossils, rocks and glaciers to demonstrate understanding of how continents were once attached. Then, they analyze data on rock layers to predict possible rock layers on continents that were once attached. Students build a model of the direction of mantle convection, identify plate boundaries and related ocean structures such as trenches and ridges and construct an explanation of plate boundaries features including sea floor structure. Finally, they analyze and interpret GPS data on current plate movement to infer numeric rates of change.

### **3.29. DYNAMIC EARTH**

Students analyze earthquake frequency at different times and model how earthquakes occur. Students determine where volcanoes, earthquakes and mountain building may occur at different types of plate boundaries, explaining how one of these might happen and estimating the range of time for these events. Students also model how crust forms on the earth.

### **3.30. PLATE TECTONICS EOU**

Students demonstrate their understanding of plate tectonics, history of the earth and earth materials and systems. In addition, students demonstrate their ability to construct explanations, analyze and interpret data and develop and use models.

### **3.31. WATER ON THE MOVE**

Students demonstrate understanding of the water cycle and energy transfer and matter cycling involved by predicting lake levels over time; they also show how energy affects water at the surface of the lake. Students model how a cloud forms over water and what happens when a cloud produces precipitation. Next, students explain what happens to water in the water cycle across a region, showing ways water cycles through the environment. Finally, students identify the driving force or energy in processes involved in the water cycle.

### **3.32. WHERE'S WATER?**

Students demonstrate their understanding of the movement and location of water underground. Students analyze maps, evaluate the risk of erosion on different types of ground, trace and model the movement of water underground and explain the water distribution in Earth's aquifers.

### **3.33. WATER AND LANDFORMS EOU**

Students demonstrate their understanding of how water moves on Earth. In addition, students demonstrate their ability to develop and use models and express cause and effect relationships.

### **3.34. WEATHER AT THE BEACH**

Students model the flow of energy that drives weather conditions on a beach. Students model the causes of the movement of air and collect data about differences in sand and water after planning an experiment.

### **3.35. STORMS**

Students model conditions that cause storms. Students model the development of clouds, the movement of air and weather conditions before and after a storm, and make predictions about weather on maps using information about air masses.

### **3.36. WEATHER EOU**

Students demonstrate their understanding of the systems and interactions that produce weather and climate; they model the warming of air by the sun and the causes of the movement of air in local settings, interpret a weather map by describing air masses and predicting weather conditions based on the data presented, model the conditions that result in seasonal differences in weather on Earth, and explain whether daylight contributes to differences in temperature in different cities.

## 4. Descriptive Statistics for All Tasks

For each task, the number of students tested, the total score frequencies and the test reliability are shown in Tables 4.1 – 4.3. Table 4.3 includes both classical raw score (Cronbach's alpha) and IRT (marginal) reliabilities. ONPAR task reports are based on raw screen and total scores. We used IRT item analysis for diagnostic use in reviewing and revising task items.

Note that six tasks include “Rescored” in the task title. Most of the tasks were revised based on the results obtained. These six tasks had minor changes to scoring rules without any changes to the screen or student response design. Therefore, we were able to recompute the student scores based on the new scoring rules and those results are reported here.

### 4.1. Number of Students Assessed

*Table 1: Number of Students Assessed for Each Task*

UnitName	TaskName	n
Body Systems	Mountain Hike	861
	Workout	866
	Body Systems EOU	821
Chemical Reactions	Science Night Rescored	1073
	Camping and Chemistry Rescored	878
	Chemical Reactions EOU Rescored	878
Ecology	Food Webs	1626
	Abiotic	1559
	Ecology EOU	1517
Energy	Energy in the world	651
	Energy in action	646
	Energy EOU	633
Heredity	Farm Genetics	1021
	Marigolds and Mosquitoes	896
	Heredity EOU	914
Light	Light and waves at the Aquarium	914
	Picture Day	896
	Light EOU	892
Matter and Energy in Living Systems	School Garden	943
	Living Things and Food	844
	Matter and Energy in Living Systems EOU	835
Motion and Forces	Space Station	1653
	Apples	1918
	Motion and Forces EOU	1649
Particle Nature of Matter	Matter in Natural Resources Rescored	1032
	Matter in the Kitchen Rescored	1015
	Particle Nature of Matter EOU Rescored	969

UnitName	TaskName	n
Plate Tectonics	Changing Continents	1062
	Dynamic Earth	1027
	Plate Tectonics EOU	1014
Water and Landforms	Water on the move	657
	Where's Water?	647
	Water and Landforms EOU	633
Weather	Weather at the Beach	2187
	Storms	2078
	Weather EOU	1925

## 4.2. Total Score Frequencies

Table 2: Total Score Frequencies for All Tasks

TaskName	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Mountain Hike	10	21	30	53	78	68	90	103	75	75	62	65	48	36	26	17	4	.	.	.	.	.	.	.	.	
Workout	20	83	94	119	146	109	96	60	49	31	18	22	7	7	1	4	.	.	.	.	.	.	.	.		
Body Systems EOU	9	16	46	52	67	73	78	88	83	66	62	56	34	31	21	15	10	11	3	.	.	.	.	.		
Science Night Rescored	14	13	34	40	59	93	130	128	101	112	98	100	72	40	24	15	.	.	.	.	.	.	.	.		
Camping and Chemistry Rescored	81	57	81	78	72	85	66	57	63	46	32	44	27	25	15	14	16	7	12	.	.	.	.	.		
Chemical Reactions EOU Rescored	8	14	40	50	54	51	54	71	74	68	59	50	46	42	39	33	38	27	22	15	23	.	.	.	.	
Food Webs	25	49	81	120	173	203	191	194	191	167	94	79	28	21	6	4	.	.	.	.	.	.	.	.		
Abiotic	24	27	60	100	118	146	165	172	165	159	139	100	76	39	37	14	9	8	1	.	.	.	.	.		
Ecology EOU	37	32	68	91	130	122	159	155	162	132	151	136	86	34	14	2	6	.	.	.	.	.	.	.		
Energy in the world	4	10	16	33	49	88	58	64	69	53	49	57	40	26	17	12	3	1	2	.	.	.	.	.		
Energy in action	5	11	27	35	67	65	77	73	53	50	61	43	37	28	6	4	4	.	.	.	.	.	.	.		
Energy EOU	2	3	12	29	20	39	38	52	46	68	45	58	35	31	39	33	23	25	17	11	4	3	.	.	.	
Farm Genetics	28	37	74	81	95	103	95	83	74	74	75	53	39	35	19	30	21	4	1	.	.	.	.	.	.	
Marigolds and Mosquitoes	72	85	110	112	144	123	97	67	44	24	9	4	2	0	1	1	1	.	.	.	.	.	.	.	.	
Heredity EOU	34	54	69	40	67	67	66	72	65	57	56	51	33	40	47	27	29	13	12	11	2	2	.	.	.	
Light and waves at the Aquarium	10	18	41	44	64	95	112	92	103	97	72	78	35	23	12	11	5	2	.	.	.	.	.	.	.	
Picture Day	23	36	67	82	81	86	95	79	74	81	56	43	30	22	19	11	7	4	.	.	.	.	.	.	.	
Light EOU	4	19	54	83	85	79	85	92	88	79	84	55	40	25	13	5	2	.	.	.	.	.	.	.	.	
School Garden	35	53	76	87	119	116	114	56	55	51	41	31	20	18	28	7	11	9	13	3	.	.	.	.	.	
Living Things and Food	9	24	44	75	106	113	108	74	102	59	42	30	24	13	5	10	4	2	.	.	.	.	.	.	.	
Matter and Energy in Living Systems EOU	14	29	52	68	62	56	76	80	75	78	60	53	43	31	26	10	11	8	3	.	.	.	.	.	.	
SpaceStation	29	56	98	143	160	156	168	162	145	125	82	89	71	85	46	29	7	1	1	.	.	.	.	.	.	
Apples	30	73	119	131	160	215	216	228	207	138	127	90	80	41	32	22	8	1	.	.	.	.	.	.	.	
Motion and Forces EOU	24	89	137	154	157	158	165	156	119	118	117	73	80	57	19	9	10	4	3	.	.	.	.	.	.	
Matter in Natural Resources Rescored	6	4	9	19	34	40	79	79	113	108	107	104	90	69	51	37	36	21	10	6	4	4	2	.	.	
Matter in the Kitchen Rescored	2	5	11	16	30	46	54	76	95	85	96	93	91	70	65	75	53	30	18	4	.	.	.	.	.	
Particle Nature of Matter EOU Rescored	7	12	37	40	64	63	96	99	93	94	85	81	65	48	47	22	9	6	1	.	.	.	.	.	.	
Changing Continents	18	26	57	59	59	64	84	94	90	72	91	67	62	69	50	40	30	14	11	3	2	.	.	.	.	.
Dynamic Earth	12	17	37	58	57	84	95	94	95	97	101	75	53	55	37	20	16	6	7	8	3	.	.	.	.	.
Plate Tectonics EOU	7	12	17	30	35	46	46	43	60	67	56	60	47	66	55	58	54	46	49	50	30	28	20	16	13	3
Water on the move	8	19	29	38	50	44	56	62	55	56	44	41	38	24	26	20	15	11	9	5	4	1	2	.	.	.
Where's Water?	2	9	16	23	42	58	73	77	74	61	57	51	33	25	15	12	12	4	3	.	.	.	.	.	.	.
Water And Landforms EOU	3	7	21	39	48	49	55	51	56	45	40	42	48	24	23	26	20	20	11	3	1	1	.	.	.	.
Weather at the Beach	37	127	210	252	299	247	227	192	151	113	102	79	64	49	16	16	4	2	.	.	.	.	.	.	.	.
Storms	42	102	146	240	316	341	295	205	167	99	63	33	21	4	2	.	2	.	.	.	.	.	.	.	.	.
Weather EOU	27	92	209	252	289	265	211	202	136	94	58	46	26	7	10	1	.	.	.	.	.	.	.	.	.	.

### 4.3. Task Reliabilities (Cronbach's Alpha and IRT)

Table 3: Raw Score (Cronbach's Alpha) and IRT Reliability for All Tasks

TaskName	Alpha	N	IRT-estimated Reliability
Mountain Hike	0.61	859	0.64
Workout	0.54	863	0.53
Body Systems EOU	0.63	819	0.65
Science Night Rescored	0.53	1042	0.49
Camping and Chemistry Rescored	0.74	839	0.59
Chemical Reactions EOU Rescored	0.76	837	0.69
Food Webs	0.62	1615	0.63
Abiotic	0.58	1551	0.58
Ecology EOU	0.70	1506	0.65
Energy in the world	0.71	648	0.75
Energy in action	0.61	645	0.59
Energy EOU	0.73	631	0.73
Farm Genetics	0.69	1019	0.65
Marigolds and Mosquitoes	0.51	894	0.44
Heredity EOU	0.72	910	0.69
Light and waves at the Aquarium	0.63	914	0.63
Picture Day	0.68	896	0.69
Light EOU	0.67	892	0.68
School Garden	0.76	925	0.69
Living Things and Food	0.68	841	0.63
Matter and Energy in Living Systems EOU	0.74	831	0.73
Space Station	0.72	1607	0.64
Apples	0.68	1907	0.65
Motion and Forces EOU	0.72	1636	0.70
Matter in Natural Resources Rescored	0.66	993	0.66
Matter in the Kitchen Rescored	0.67	979	0.66
Particle Nature of Matter EOU Rescored	0.66	941	0.62
Changing Continents	0.72	1061	0.64
Dynamic Earth	0.72	1025	0.68
Plate Tectonics EOU	0.81	1013	0.79
Water on the move	0.67	657	0.64
Where's Water?	0.63	647	0.64
Water And Landforms EOU	0.75	633	0.76
Weather at the Beach	0.64	2183	0.64
Storms	0.52	2018	0.50
Weather EOU	0.53	1923	0.51

## 5. Reliability and Item Statistics for Each Task

In this section, a variety of statistics were computed for each task and the results are produced in four tables and two plots. The first table shows the frequency of the responses to each item. We want to see that all possible responses are being used across our student sample.

The second table shows the correlation between each item score and the total score for the task. We look for correlations above .10. The table column “Alpha.Without” shows the Cronbach’s alpha for the test with that item removed. For this statistic, we look for values that are lower than the alpha for the full test. If a task is just as reliable without the item, that item may not be adding information to the task.

The Item Step Locations table shows the IRT difficulties of each step of the polytomous items. This information is related to the item response frequency and again we look for an even distribution differences between steps. In addition, we can see if the steps across all task items cover the range from -2 to +2. (The items were calibrated such that the mean student score is zero.)

The last table in this section shows item fit. This is an IRT statistic that indicates how well each item matches the other items in the task. Here we are using Orlando and Thissen (2000, 2003) and Kang and Chen’s (2007) signed chi-squared test. Low fit suggests the item may not be contributing useful information to the total task score. A probability value below .01 is a criterion for item review.

The two plots display the task reliability as a function of student ability (Theta on the x-axis). Here we look for tasks that show high test information or reliability and low error of measurement across a reasonable range for student ability. A task that is only reliable for students of high ability (shifted right) or low ability (shifted left) might need to be revised to better assess all students.

### 5.1. Mountain Hike

*Table 4: Item Response Frequencies*

Item	0	1	2	3	4
s_12_57_02	166	343	177	173	
s_12_57_03	415	163	281		
s_12_57_05	476	78	134	171	
s_12_57_06	74	555	187	43	
s_12_57_08	689	119	51		
s_12_57_10	156	703			
s_12_57_11	112	255	188	157	147

*Table 5: Item-Total Correlations and Alpha Without*

Variable	Item.Total	Alpha.Without	N
s_12_57_02	0.442	0.534	859
s_12_57_03	0.398	0.554	859
s_12_57_05	0.211	0.637	859
s_12_57_06	0.349	0.577	859
s_12_57_08	0.286	0.595	859
s_12_57_10	0.262	0.606	859
s_12_57_11	0.492	0.509	859

*Table 6: Item Step Locations*

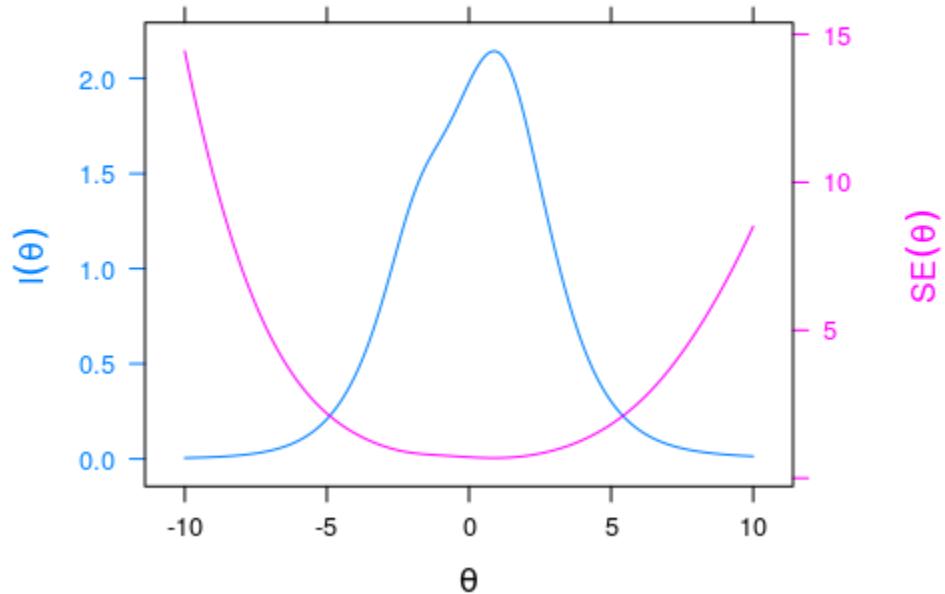
Item	nCats	Step_1	Step_2	Step_3	Step_4
s_12_57_02	3.0000	-1.5538	0.2735	1.6341	
s_12_57_03	2.0000	-0.3324	1.0553		
s_12_57_05	3.0000	0.2529	1.6180	3.6107	
s_12_57_06	3.0000	-3.0923	1.2925	3.8220	
s_12_57_08	2.0000	1.2104	2.8865		
s_12_57_10	1.0000	-1.9547			
s_12_57_11	4.0000	-2.0739	-0.4295	0.6462	1.8650

*Table 7: IRT Item Fit (S-X2)*

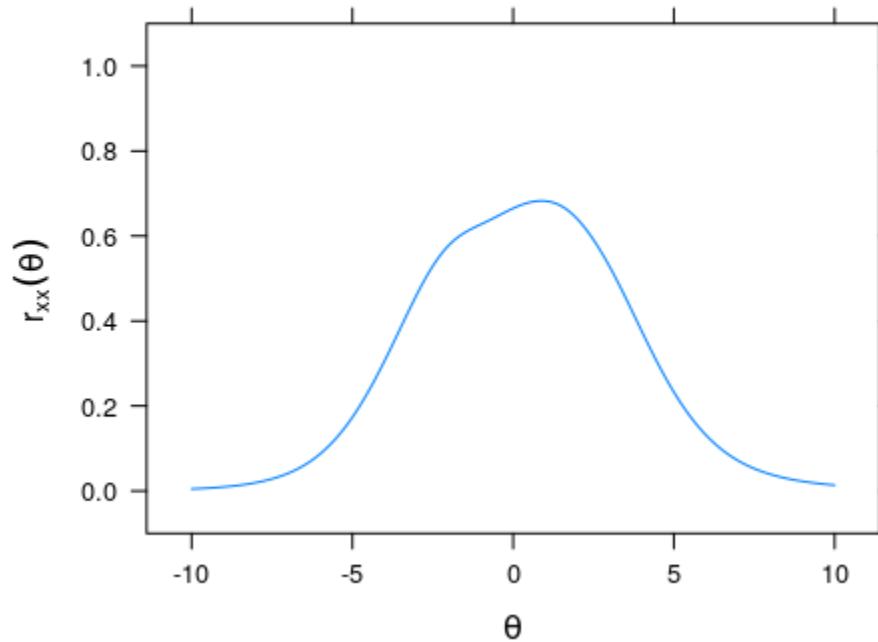
item	Fit	df	Probability
s_12_57_02	29.712	30	0.480
s_12_57_03	54.119	23	0.000
s_12_57_05	59.168	35	0.007
s_12_57_06	42.575	30	0.064
s_12_57_08	31.193	22	0.092
s_12_57_10	6.062	12	0.913
s_12_57_11	28.658	34	0.727

Probability < .01 indicates poor category fit

**Test Information (blue) and Standard Error (red)  
by Student Ability**



**Reliability by Student Ability**



## 5.2. Workout

*Table 8: Item Response Frequencies*

Item	0	1	2	3	4
s_12_58_02	58	354	271	181	
s_12_58_03	714	16	114	20	
s_12_58_04	413	327	121	3	
s_12_58_05	238	215	273	137	
s_12_58_07	796	30	19	10	9
s_12_58_08	661	43	63	97	

*Table 9: Item-Total Correlations and Alpha Without*

Variable	Item.Total	Alpha.Without	N
s_12_58_02	0.254	0.507	863
s_12_58_03	0.353	0.463	863
s_12_58_04	0.306	0.486	863
s_12_58_05	0.284	0.498	863
s_12_58_07	0.241	0.514	863
s_12_58_08	0.301	0.487	863

*Table 10: Item Step Locations*

Item	nCats	Step_1	Step_2	Step_3	Step_4
s_12_58_02	3.0000	-3.1791	-0.1990	1.7645	
s_12_58_03	3.0000	1.1224	2.0210	3.6642	
s_12_58_04	3.0000	-0.1788	2.7361	8.0515	
s_12_58_05	3.0000	-1.3395	0.2375	2.1119	
s_12_58_07	4.0000	1.5748	2.3752	2.9830	3.8165
s_12_58_08	3.0000	0.8534	1.6925	2.6298	

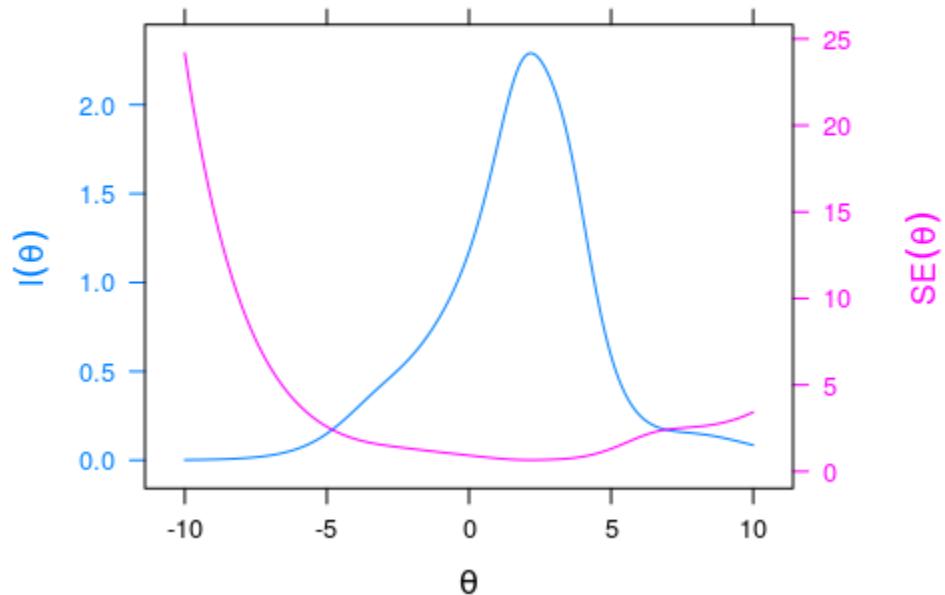
*Table 11: IRT Item Fit (S-X2)*

item	Fit	df	Probability
s_12_58_02	30.972	23	0.123
s_12_58_03	43.276	25	0.013
s_12_58_04	16.424	19	0.629
s_12_58_05	35.501	26	0.101
s_12_58_07	44.661	29	0.032
s_12_58_08	40.060	28	0.065

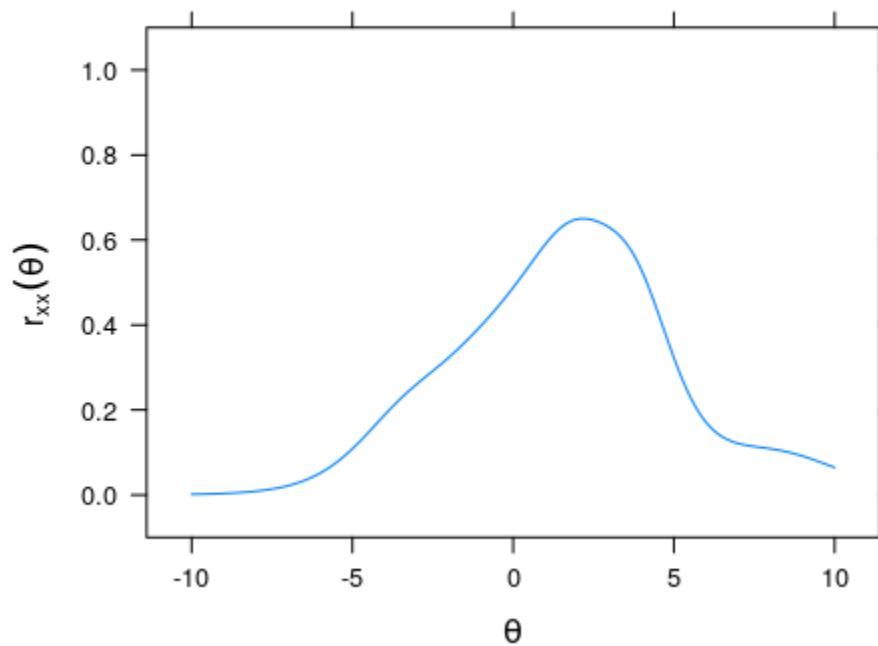
Probability < .01 indicates poor category fit

---

**Test Information (blue) and Standard Error (red)  
by Student Ability**



**Reliability by Student Ability**



### 5.3. Body Systems EOU

*Table 12: Item Response Frequencies*

Item	0	1	2	3
s_12_59_02	289	478	52	
s_12_59_04	412	321	74	12
s_12_59_05	371	347	29	72
s_12_59_07	227	188	183	221
s_12_59_08	69	455	214	81
s_12_59_09	237	222	242	118
s_12_59_11	485	168	165	1
s_12_59_12	530	87	112	90

*Table 13: Item-Total Correlations and Alpha Without*

Variable	Item.Total	Alpha.Without	N
s_12_59_02	0.324	0.603	819
s_12_59_04	0.385	0.585	819
s_12_59_05	0.262	0.613	819
s_12_59_07	0.426	0.563	819
s_12_59_08	0.529	0.546	819
s_12_59_09	0.363	0.585	819
s_12_59_11	0.083	0.653	819
s_12_59_12	0.297	0.607	819

*Table 14: Item Step Locations*

Item	nCats	Step_1	Step_2	Step_3
s_12_59_02	2.0000	-1.2900	5.6093	
s_12_59_04	3.0000	-0.0485	2.3149	4.4670
s_12_59_05	3.0000	-0.5311	4.5655	6.0345
s_12_59_07	3.0000	-1.2088	0.0509	1.2997
s_12_59_08	3.0000	-2.2086	0.4519	2.1920
s_12_59_09	3.0000	-1.2664	0.4626	2.4453
s_12_59_11	3.0000	3.4841	10.0000	10.0000
s_12_59_12	3.0000	0.5829	1.7249	3.1358

*Table 15: IRT Item Fit (S-X2)*

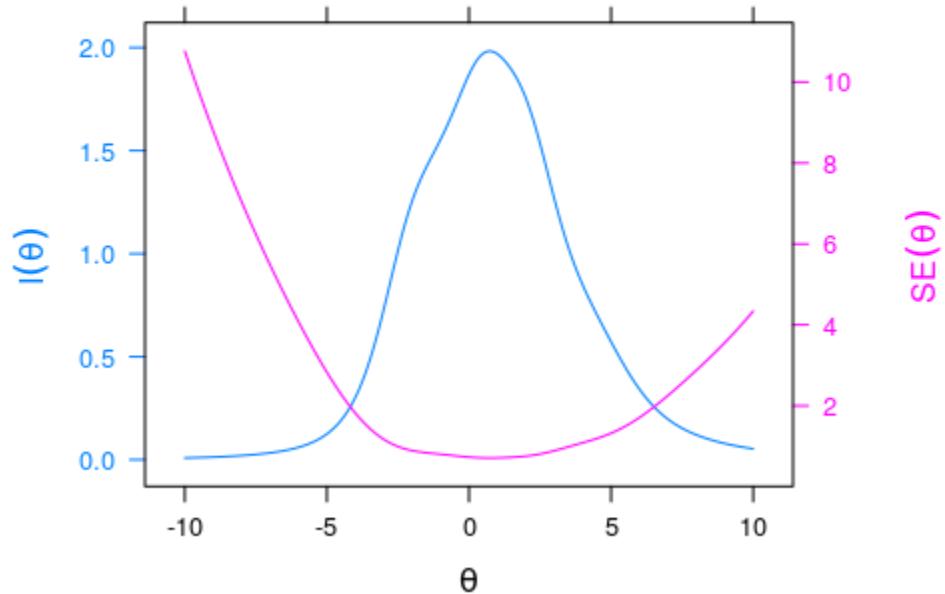
item	Fit	df	Probability
s_12_59_02	56.076	26	0.001

item	Fit	df	Probability
s_12_59_04	34.753	32	0.338
s_12_59_05	73.549	39	0.001
s_12_59_07	29.859	33	0.624
s_12_59_08	38.106	31	0.178
s_12_59_09	75.919	35	0.000
s_12_59_11	46.803	29	0.019
s_12_59_12	36.688	38	0.530

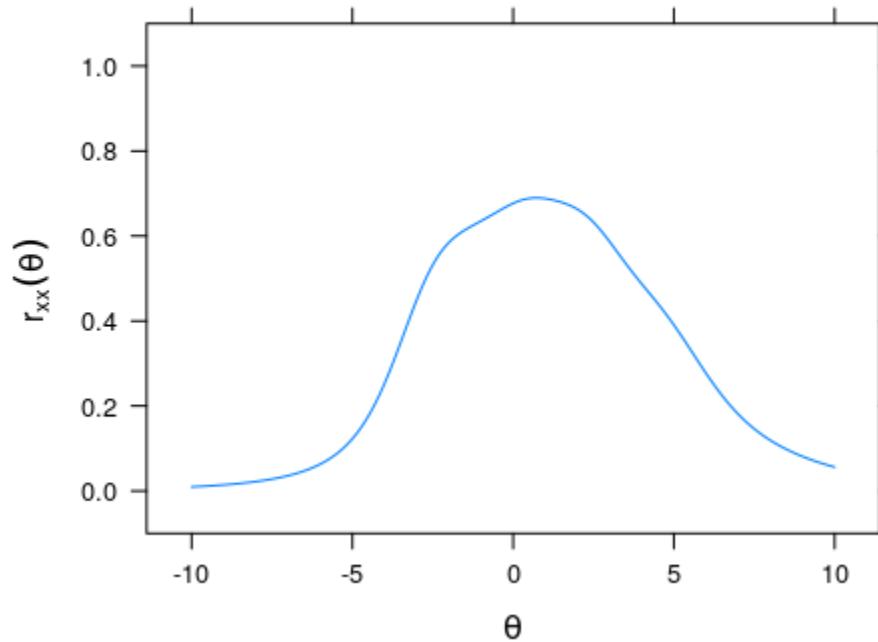
Probability < .01 indicates poor category fit

---

**Test Information (blue) and Standard Error (red)  
by Student Ability**



**Reliability by Student Ability**



## 5.4. Science Night Rescored

*Table 16: Item Response Frequencies*

Item	0	1	2	3
s_09_44_03	314	203	430	95
s_09_44_04	400	134	95	413
s_09_44_06	73	279	467	223
s_09_44_07	89	93	371	489
s_09_44_09	481	371	219	
s_09_44_10	618	345	79	

*Table 17: Item-Total Correlations and Alpha Without*

Variable	Item.Total	Alpha.Without	N
s_09_44_03	0.279	0.483	1042
s_09_44_04	0.384	0.426	1042
s_09_44_06	0.193	0.521	1042
s_09_44_07	0.267	0.489	1042
s_09_44_09	0.299	0.479	1042
s_09_44_10	0.291	0.489	1042

*Table 18: Item Step Locations*

Item	nCats	Step_1	Step_2	Step_3
s_09_44_03	3.0000	-1.3950	0.1238	3.0011
s_09_44_04	3.0000	-0.9616	-0.0195	0.8648
s_09_44_06	3.0000	-9.0812	-2.3721	4.4510
s_09_44_07	3.0000	-4.6830	-2.8505	0.2394
s_09_44_09	2.0000	-0.4810	2.5108	
s_09_44_10	2.0000	0.5303	3.9822	

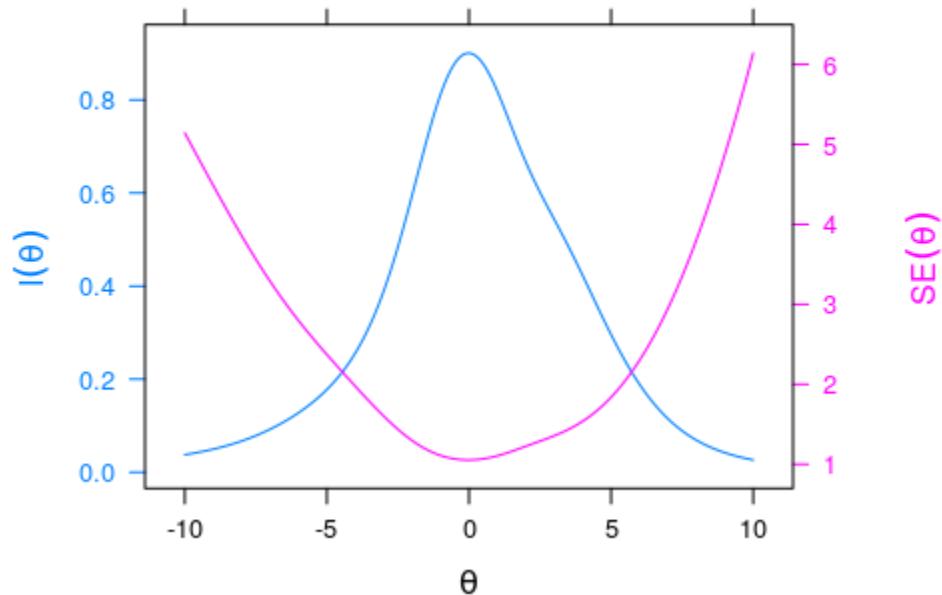
*Table 19: IRT Item Fit (S-X2)*

item	Fit	df	Probability
s_09_44_03	64.427	27	0.000
s_09_44_04	37.943	26	0.061
s_09_44_06	36.928	29	0.148
s_09_44_07	41.000	28	0.054
s_09_44_09	48.421	21	0.001
s_09_44_10	45.388	21	0.002

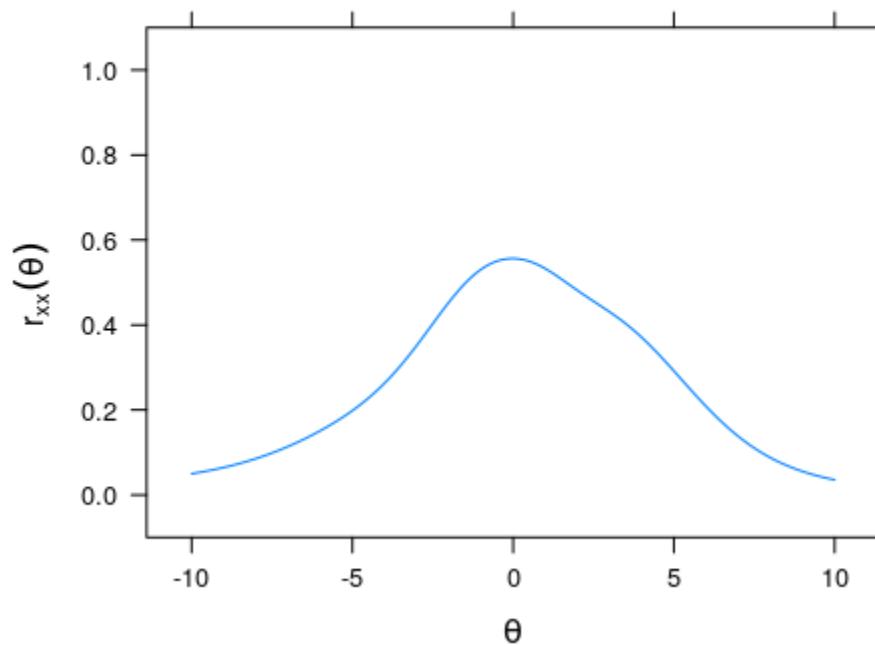
Probability < .01 indicates poor category fit

---

**Test Information (blue) and Standard Error (red)  
by Student Ability**



**Reliability by Student Ability**



## 5.5. Camping and Chemistry Rescored

*Table 20: Item Response Frequencies*

Item	0	1	2	3	4
s_09_45_03	450	74	315		
s_09_45_04	413	196	132	98	
s_09_45_06	322	152	140	138	124
s_09_45_08	660	21	64	37	57
s_09_45_09	361	145	370		
s_09_45_11	298	246	78	254	

*Table 21: Item-Total Correlations and Alpha Without*

Variable	Item.Total	Alpha.Without	N
s_09_45_03	0.256	0.758	839
s_09_45_04	0.575	0.681	839
s_09_45_06	0.568	0.681	839
s_09_45_08	0.491	0.703	839
s_09_45_09	0.500	0.705	839
s_09_45_11	0.518	0.695	839

*Table 22: Item Step Locations*

Item	nCats	Step_1	Step_2	Step_3	Step_4
s_09_45_03	2.0000	-0.0597	1.2029		
s_09_45_04	3.0000	-0.3027	0.8697	2.0447	
s_09_45_06	4.0000	-0.8142	0.1563	0.9325	2.0212
s_09_45_08	4.0000	0.5810	1.2506	1.8162	2.6075
s_09_45_09	2.0000	-0.6133	0.6627		
s_09_45_11	3.0000	-0.8029	0.3384	1.2487	

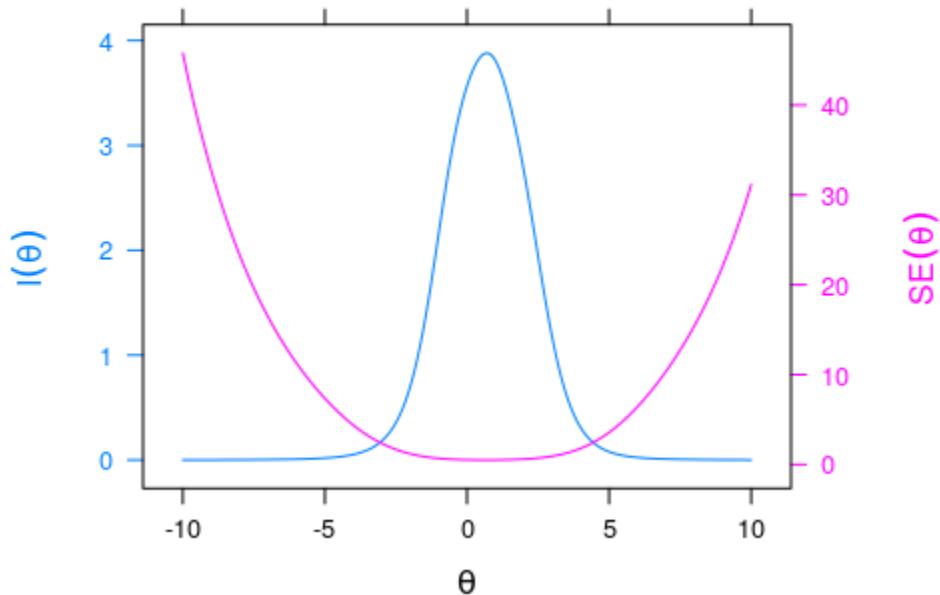
*Table 23: IRT Item Fit (S-X2)*

item	Fit	df	Probability
s_09_45_03	60.736	28	0.000
s_09_45_04	36.453	34	0.355
s_09_45_06	54.074	40	0.068
s_09_45_08	34.217	38	0.645
s_09_45_09	44.676	25	0.009
s_09_45_11	25.369	33	0.826

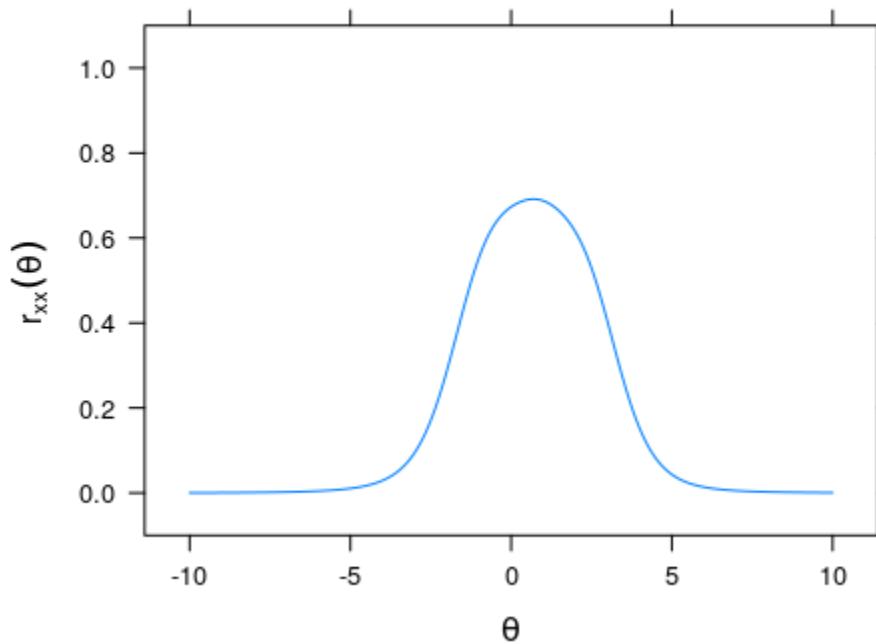
Probability < .01 indicates poor category fit

---

**Test Information (blue) and Standard Error (red)  
by Student Ability**



**Reliability by Student Ability**



## 5.6. Chemical Reactions EOU Rescored

*Table 24: Item Response Frequencies*

Item	0	1	2	3	4
s_09_47_03	207	249	420		
s_09_47_04	353	192	330		
s_09_47_05	517	104	46	209	
s_09_47_07	75	335	201	120	145
s_09_47_08	84	137	185	470	
s_09_47_10	513	68	107	36	114
s_09_47_11	333	74	469		

*Table 25: Item-Total Correlations and Alpha Without*

Variable	Item.Total	Alpha.Without	N
s_09_47_03	0.387	0.750	837
s_09_47_04	0.433	0.742	837
s_09_47_05	0.511	0.725	837
s_09_47_07	0.632	0.695	837
s_09_47_08	0.435	0.740	837
s_09_47_10	0.504	0.732	837
s_09_47_11	0.495	0.730	837

*Table 26: Item Step Locations*

Item	nCats	Step_1	Step_2	Step_3	Step_4
s_09_47_03	2.0000	-1.5435	0.3406		
s_09_47_04	2.0000	-0.6716	0.8400		
s_09_47_05	3.0000	-0.0372	0.8223	1.6129	
s_09_47_07	4.0000	-1.8996	-0.2431	0.7075	1.7211
s_09_47_08	3.0000	-2.3266	-1.0187	0.2196	
s_09_47_10	4.0000	-0.0598	0.6992	1.3070	2.0480
s_09_47_11	2.0000	-0.7771	0.3295		

*Table 27: IRT Item Fit (S-X2)*

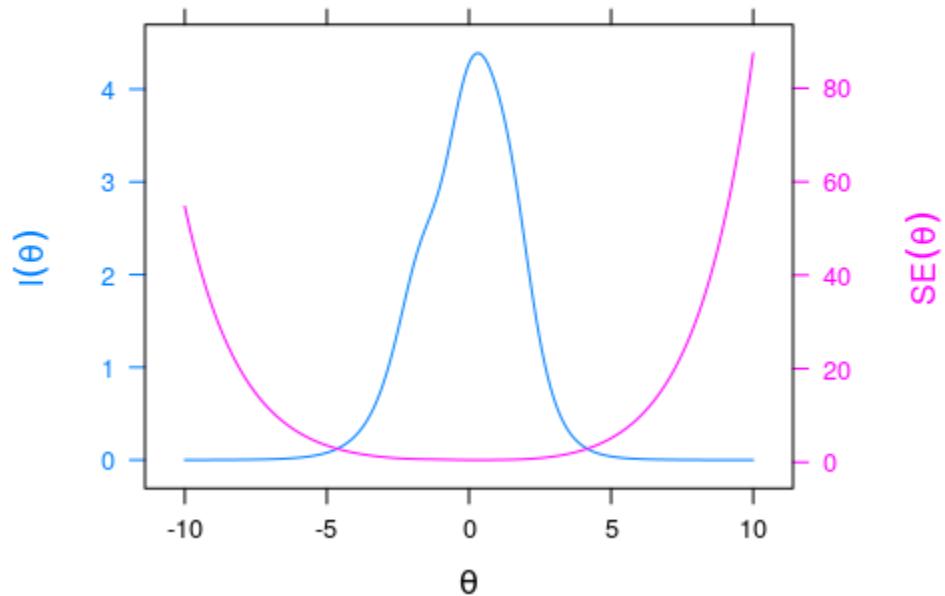
item	Fit	df	Probability
s_09_47_03	25.444	31	0.748
s_09_47_04	60.223	31	0.001
s_09_47_05	35.522	39	0.629

item	Fit	df	Probability
s_09_47_07	37.245	40	0.595
s_09_47_08	33.391	35	0.546
s_09_47_10	46.546	43	0.329
s_09_47_11	45.634	29	0.026

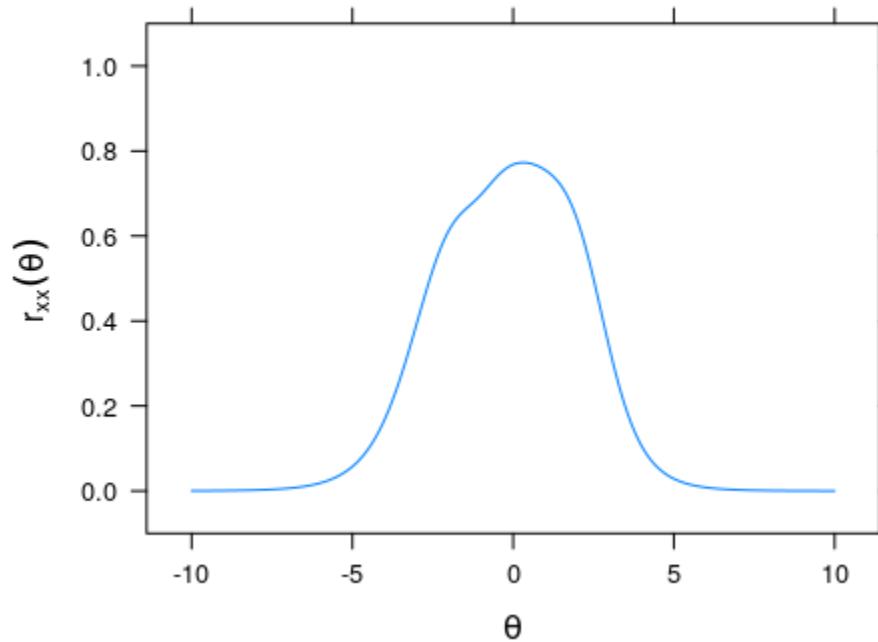
Probability < .01 indicates poor category fit

---

**Test Information (blue) and Standard Error (red)  
by Student Ability**



**Reliability by Student Ability**



## 5.7. Food Webs

*Table 28: Item Response Frequencies*

Item	0	1	2	3
s_06_22_02	249	917	379	77
s_06_22_03	712	504	405	
s_06_22_05	875	747		
s_06_22_06	488	416	717	
s_06_22_07	842	612	167	
s_06_22_10	770	800	48	
s_06_22_11	536	1081	5	
s_06_22_12	642	357	623	

*Table 29: Item-Total Correlations and Alpha Without*

Variable	Item.Total	Alpha.Without	N
s_06_22_02	0.448	0.541	1615
s_06_22_03	0.360	0.569	1615
s_06_22_05	0.231	0.604	1615
s_06_22_06	0.296	0.592	1615
s_06_22_07	0.442	0.547	1615
s_06_22_10	0.311	0.586	1615
s_06_22_11	0.232	0.604	1615
s_06_22_12	0.233	0.616	1615

*Table 30: Item Step Locations*

Item	nCats	Step_1	Step_2	Step_3
s_06_22_02	3.0000	-1.6957	0.8397	2.7870
s_06_22_03	2.0000	-0.4628	1.4141	
s_06_22_05	1.0000	0.2734		
s_06_22_06	2.0000	-1.4106	0.4461	
s_06_22_07	2.0000	-0.0536	1.9968	
s_06_22_10	2.0000	-0.1772	5.3400	
s_06_22_11	2.0000	-1.7841	6.0000	
s_06_22_12	2.0000	-1.1811	1.2831	

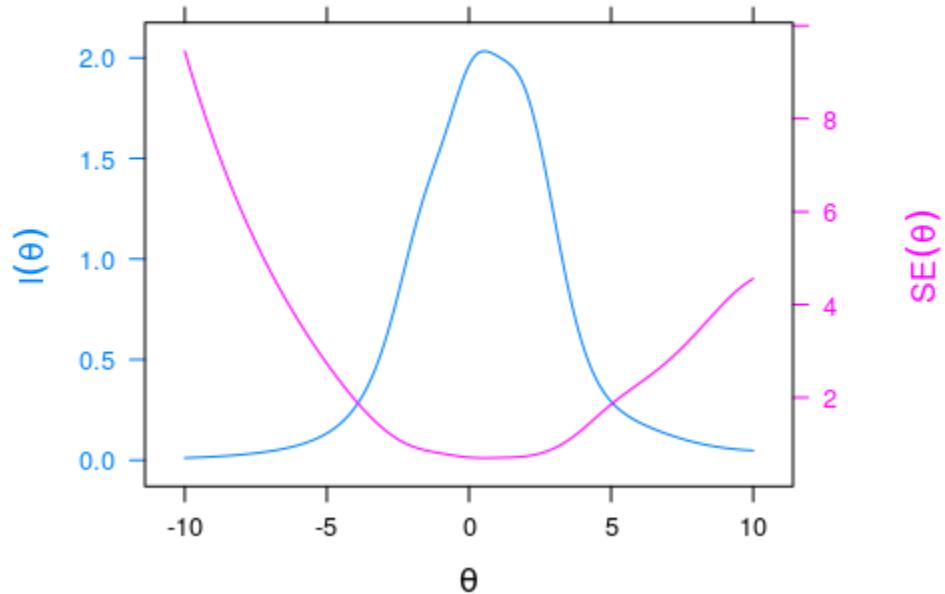
*Table 31: IRT Item Fit (S-X2)*

item	Fit	df	Probability
s_06_22_02	23.253	24	0.505
s_06_22_03	16.329	19	0.635
s_06_22_05	6.356	11	0.849
s_06_22_06	37.038	20	0.012
s_06_22_07	16.465	18	0.560
s_06_22_10	25.010	21	0.247
s_06_22_11	13.638	10	0.190
s_06_22_12	29.416	21	0.104

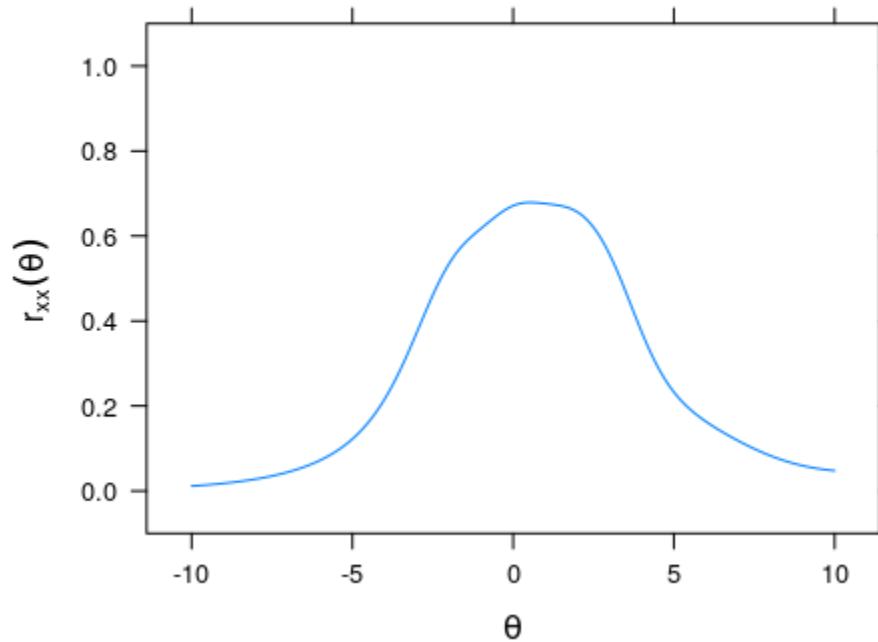
Probability < .01 indicates poor category fit

---

**Test Information (blue) and Standard Error (red)  
by Student Ability**



**Reliability by Student Ability**



## 5.8. Abiotic

*Table 32: Item Response Frequencies*

Item	0	1	2	3
s_06_24_02	226	391	937	
s_06_24_03	514	274	222	544
s_06_24_04	506	369	440	238
s_06_24_06	724	588	166	77
s_06_24_07	322	633	600	
s_06_24_09	588	652	315	
s_06_24_10	1156	256	24	119

*Table 33: Item-Total Correlations and Alpha Without*

Variable	Item.Total	Alpha.Without	N
s_06_24_02	0.351	0.529	1551
s_06_24_03	0.415	0.493	1551
s_06_24_04	0.164	0.600	1551
s_06_24_06	0.288	0.546	1551
s_06_24_07	0.333	0.534	1551
s_06_24_09	0.340	0.532	1551
s_06_24_10	0.283	0.547	1551

*Table 34: Item Step Locations*

Item	nCats	Step_1	Step_2	Step_3
s_06_24_02	2.0000	-2.0965	-0.3108	
s_06_24_03	3.0000	-1.0380	-0.0126	0.9539
s_06_24_04	3.0000	-2.4900	0.7765	5.4532
s_06_24_06	3.0000	-0.3743	2.6720	4.9581
s_06_24_07	2.0000	-1.8160	0.6441	
s_06_24_09	2.0000	-0.7217	1.8311	
s_06_24_10	3.0000	1.0928	2.5500	3.4692

*Table 35: IRT Item Fit (S-X2)*

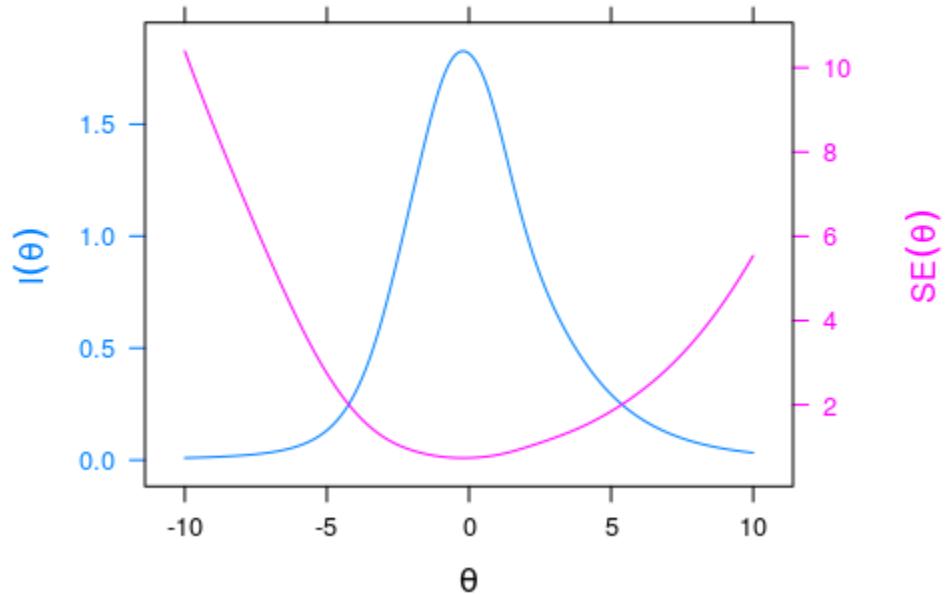
item	Fit	df	Probability
s_06_24_02	47.354	21	0.001
s_06_24_03	42.836	31	0.077
s_06_24_04	54.401	36	0.025
s_06_24_06	86.372	34	0.000

item	Fit	df	Probability
s_06_24_07	30.609	24	0.165
s_06_24_09	29.720	25	0.235
s_06_24_10	54.397	31	0.006

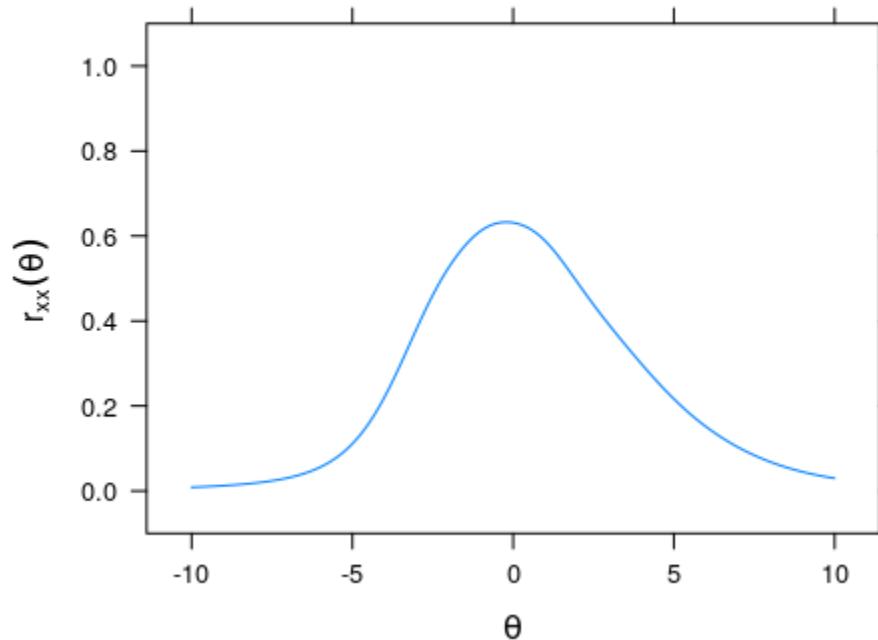
Probability < .01 indicates poor category fit

---

**Test Information (blue) and Standard Error (red)  
by Student Ability**



**Reliability by Student Ability**



## 5.9. Ecology EOU

*Table 36: Item Response Frequencies*

Item	0	1	2	3
s_06_25_02	192	273	1047	
s_06_25_04	448	662	398	
s_06_25_06	699	467	346	
s_06_25_07	564	277	672	
s_06_25_08	411	361	741	
s_06_25_09	1055	458		
s_06_25_10	880	513	120	
s_06_25_11	473	887	130	23

*Table 37: Item-Total Correlations and Alpha Without*

Variable	Item.Total	Alpha.Without	N
s_06_25_02	0.410	0.669	1506
s_06_25_04	0.324	0.689	1506
s_06_25_06	0.402	0.671	1506
s_06_25_07	0.565	0.627	1506
s_06_25_08	0.404	0.671	1506
s_06_25_09	0.324	0.689	1506
s_06_25_10	0.474	0.658	1506
s_06_25_11	0.256	0.700	1506

*Table 38: Item Step Locations*

Item	nCats	Step_1	Step_2	Step_3
s_06_25_02	2.0000	-2.0891	-0.5709	
s_06_25_04	2.0000	-1.4730	1.5950	
s_06_25_06	2.0000	-0.3651	1.5006	
s_06_25_07	2.0000	-0.7634	0.4992	
s_06_25_08	2.0000	-1.3803	0.2570	
s_06_25_09	1.0000	1.1340		
s_06_25_10	2.0000	0.1508	2.1862	
s_06_25_11	3.0000	-1.3317	3.4361	6.0000

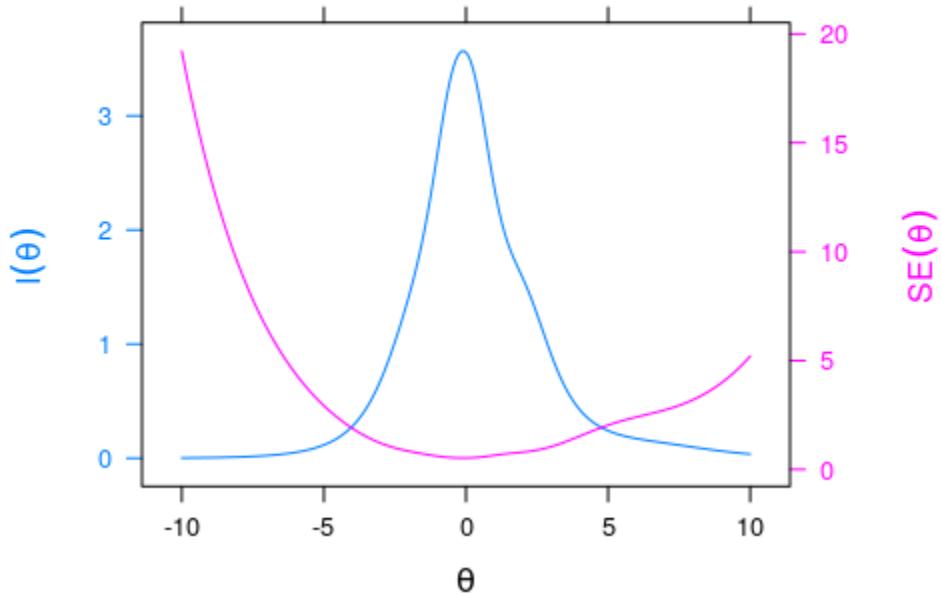
*Table 39: IRT Item Fit (S-X2)*

<b>item</b>	<b>Fit</b>	<b>df</b>	<b>Probability</b>
s_06_25_02	28.449	18	0.056
s_06_25_04	21.719	22	0.477
s_06_25_06	19.107	21	0.578
s_06_25_07	13.834	18	0.740
s_06_25_08	26.980	21	0.172
s_06_25_09	14.583	12	0.265
s_06_25_10	32.703	19	0.026
s_06_25_11	37.571	26	0.066

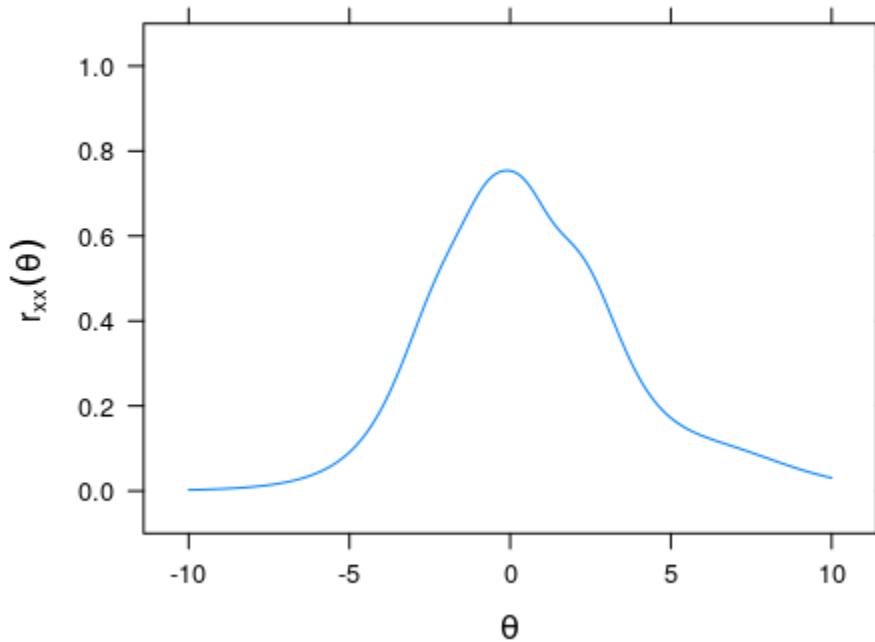
Probability < .01 indicates poor category fit

---

**Test Information (blue) and Standard Error (red)  
by Student Ability**



**Reliability by Student Ability**



## 5.10. Energy in the world

*Table 40: Item Response Frequencies*

Item	0	1	2	3
s_16_73_02	20	151	480	
s_16_73_03	116	472	60	
s_16_73_05	361	178	112	
s_16_73_06	86	336	229	
s_16_73_07	75	453	41	82
s_16_73_08	240	117	141	153
s_16_73_10	454	136	61	
s_16_73_11	435	172	44	

*Table 41: Item-Total Correlations and Alpha Without*

Variable	Item.Total	Alpha.Without	N
s_16_73_02	0.345	0.699	648
s_16_73_03	0.457	0.683	648
s_16_73_05	0.386	0.690	648
s_16_73_06	0.397	0.688	648
s_16_73_07	0.321	0.705	648
s_16_73_08	0.577	0.651	648
s_16_73_10	0.417	0.685	648
s_16_73_11	0.470	0.676	648

*Table 42: Item Step Locations*

Item	nCats	Step_1	Step_2	Step_3
s_16_73_02	2.0000	-3.7938	-1.1032	
s_16_73_03	2.0000	-1.4302	2.0691	
s_16_73_05	2.0000	0.1320	2.0564	
s_16_73_06	2.0000	-2.4201	0.8225	
s_16_73_07	3.0000	-2.9723	1.8507	3.1524
s_16_73_08	3.0000	-0.8220	0.2165	1.3311
s_16_73_10	2.0000	0.7020	2.4763	
s_16_73_11	2.0000	0.4738	2.3724	

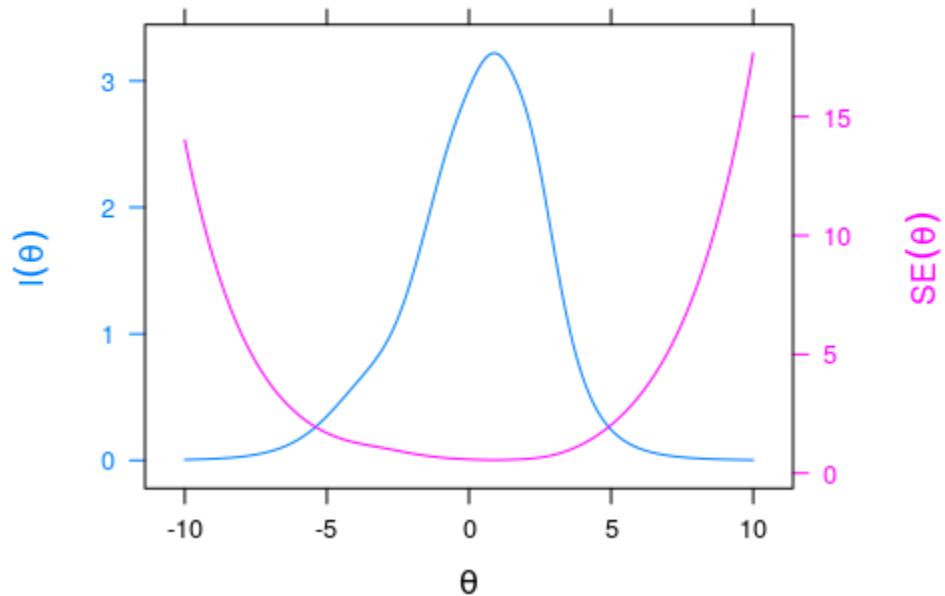
*Table 43: IRT Item Fit (S-X2)*

item	Fit	df	Probability
s_16_73_02	31.656	15	0.007
s_16_73_03	29.750	19	0.055
s_16_73_05	39.664	21	0.008
s_16_73_06	35.498	21	0.025
s_16_73_07	44.441	30	0.043
s_16_73_08	26.726	25	0.370
s_16_73_10	31.801	21	0.061
s_16_73_11	14.051	18	0.726

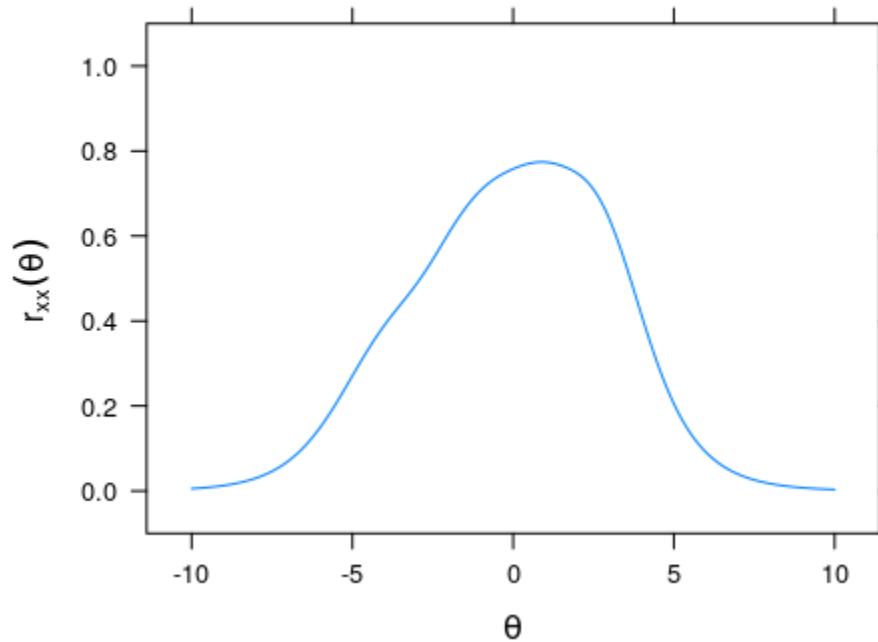
Probability < .01 indicates poor category fit

---

**Test Information (blue) and Standard Error (red)  
by Student Ability**



**Reliability by Student Ability**



## 5.11. Energy in action

*Table 44: Item Response Frequencies*

Item	0	1	2	3	4
s_16_74_02	160	122	119	245	
s_16_74_04	182	214	174	60	16
s_16_74_06	296	122	112	115	
s_16_74_07	117	240	94	195	
s_16_74_09	51	132	425	30	8

*Table 45: Item-Total Correlations and Alpha Without*

Variable	Item.Total	Alpha.Without	N
s_16_74_02	0.402	0.529	645
s_16_74_04	0.450	0.504	645
s_16_74_06	0.306	0.583	645
s_16_74_07	0.318	0.574	645
s_16_74_09	0.368	0.561	645

*Table 46: Item Step Locations*

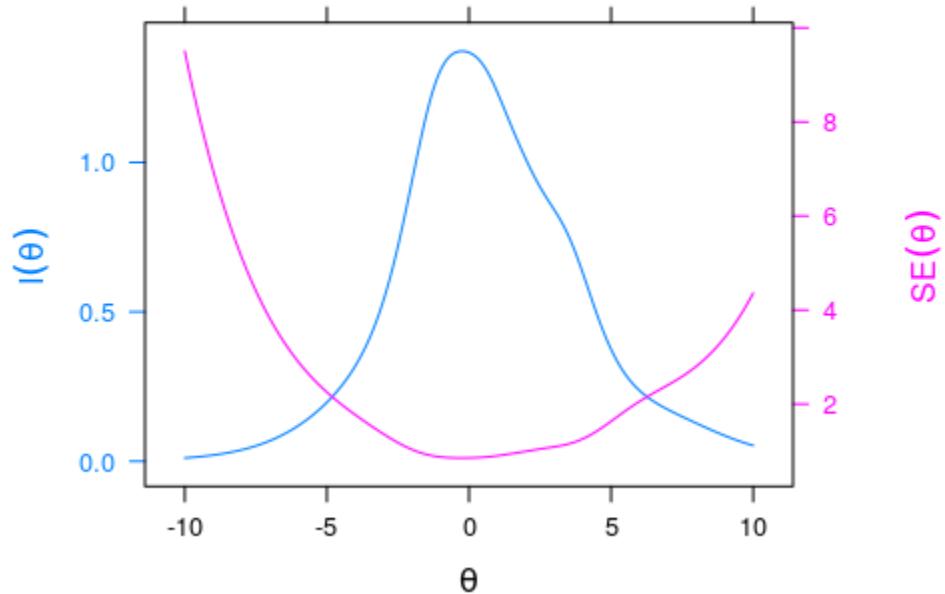
Item	nCats	Step_1	Step_2	Step_3	Step_4
s_16_74_02	3.0000	-1.4546	-0.3022	0.8186	
s_16_74_04	4.0000	-1.1225	0.4784	1.9503	3.5347
s_16_74_06	3.0000	-0.5417	1.4038	3.5734	
s_16_74_07	3.0000	-2.2524	0.1113	1.4883	
s_16_74_09	4.0000	-3.9692	-1.4183	4.2913	6.8700

*Table 47: IRT Item Fit (S-X2)*

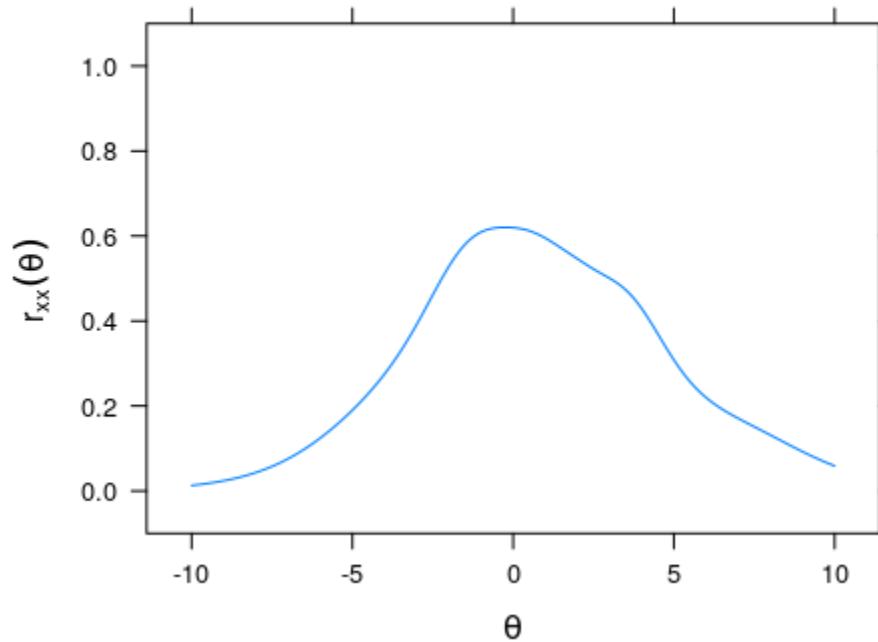
item	Fit	df	Probability
s_16_74_02	44.712	26	0.013
s_16_74_04	41.541	27	0.037
s_16_74_06	28.937	30	0.521
s_16_74_07	42.910	27	0.027
s_16_74_09	36.191	29	0.168

Probability < .01 indicates poor category fit

**Test Information (blue) and Standard Error (red)  
by Student Ability**



**Reliability by Student Ability**



## 5.12. Energy EOU

*Table 48: Item Response Frequencies*

Item	0	1	2	3
s_16_77_02	14	341	278	
s_16_77_03	100	264	219	48
s_16_77_05	167	200	161	105
s_16_77_06	213	120	246	54
s_16_77_08	179	127	155	172
s_16_77_09	201	199	111	122
s_16_77_11	162	214	173	84
s_16_77_12	264	280	89	

*Table 49: Item-Total Correlations and Alpha Without*

Variable	Item.Total	Alpha.Without	N
s_16_77_02	0.476	0.702	631
s_16_77_03	0.497	0.685	631
s_16_77_05	0.390	0.706	631
s_16_77_06	0.452	0.692	631
s_16_77_08	0.451	0.694	631
s_16_77_09	0.411	0.702	631
s_16_77_11	0.402	0.702	631
s_16_77_12	0.381	0.708	631

*Table 50: Item Step Locations*

Item	nCats	Step_1	Step_2	Step_3
s_16_77_02	2.0000	-3.3405	0.2884	
s_16_77_03	3.0000	-1.6803	0.3474	2.4330
s_16_77_05	3.0000	-1.5145	0.3135	2.1604
s_16_77_06	3.0000	-1.0321	0.2734	2.4973
s_16_77_08	3.0000	-1.2386	0.0130	1.3457
s_16_77_09	3.0000	-1.0962	0.6021	2.0075
s_16_77_11	3.0000	-1.5220	0.5172	2.5844
s_16_77_12	2.0000	-0.4555	2.3755	

*Table 51: IRT Item Fit (S-X2)*

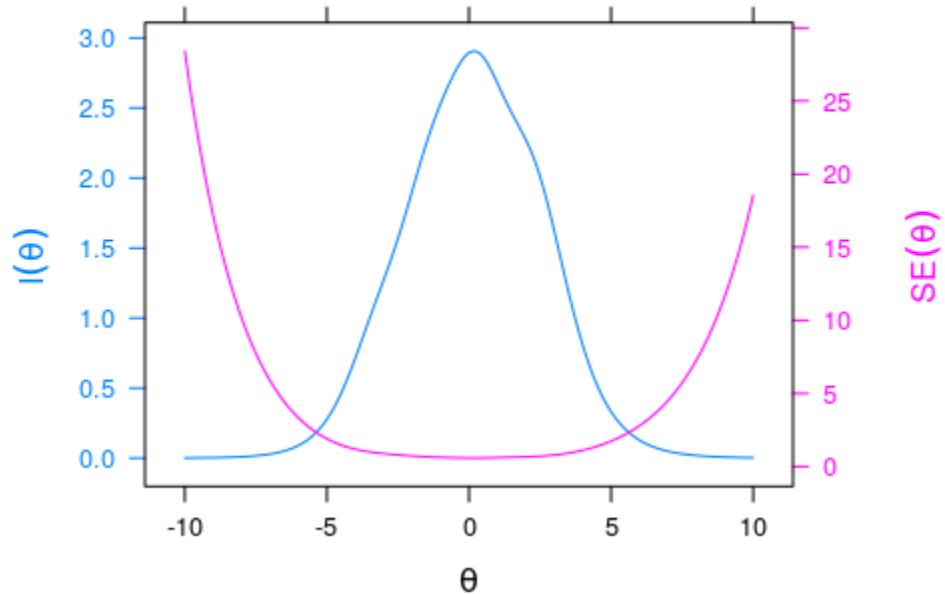
item	Fit	df	Probability
s_16_77_02	24.358	20	0.227

item	Fit	df	Probability
s_16_77_03	27.608	35	0.809
s_16_77_05	75.163	41	0.001
s_16_77_06	57.777	36	0.012
s_16_77_08	28.587	39	0.890
s_16_77_09	37.825	41	0.613
s_16_77_11	29.406	41	0.912
s_16_77_12	31.307	31	0.451

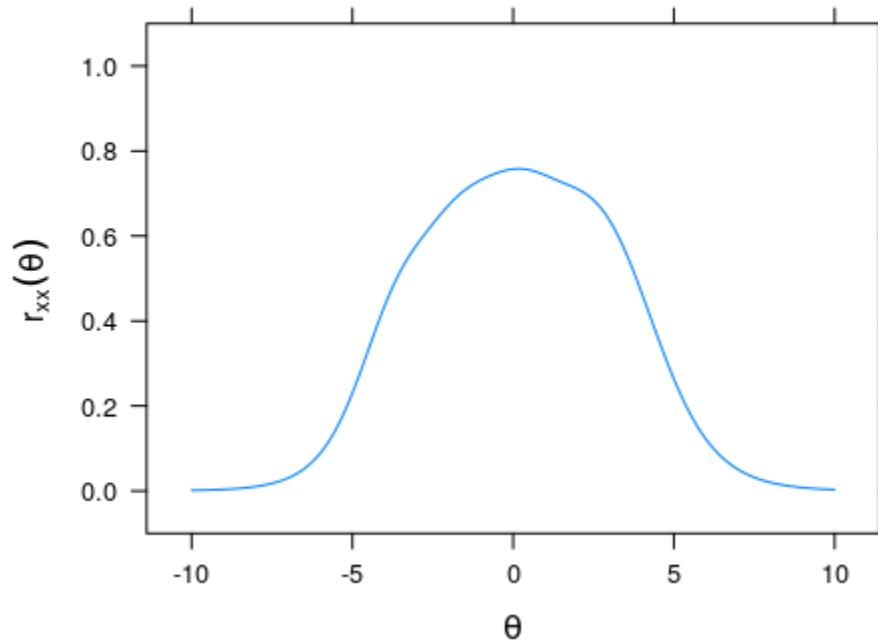
Probability < .01 indicates poor category fit

---

**Test Information (blue) and Standard Error (red)  
by Student Ability**



**Reliability by Student Ability**



## 5.13. Farm Genetics

*Table 52: Item Response Frequencies*

Item	0	1	2	3	4
s_11_53_03	323	39	657		
s_11_53_04	253	241	525		
s_11_53_06	395	300	200	110	14
s_11_53_07	370	269	23	269	88
s_11_53_09	455	242	172	150	
s_11_53_10	547	87	367	18	

*Table 53: Item-Total Correlations and Alpha Without*

Variable	Item.Total	Alpha.Without	N
s_11_53_03	0.340	0.672	1019
s_11_53_04	0.414	0.655	1019
s_11_53_06	0.436	0.643	1019
s_11_53_07	0.464	0.642	1019
s_11_53_09	0.462	0.634	1019
s_11_53_10	0.439	0.644	1019

*Table 54: Item Step Locations*

Item	nCats	Step_1	Step_2	Step_3	Step_4
s_11_53_03	2.0000	-1.3572	-0.3046		
s_11_53_04	2.0000	-1.3911	0.1303		
s_11_53_06	4.0000	-0.7433	0.7428	2.1057	4.1459
s_11_53_07	4.0000	-0.9481	0.2771	1.0035	2.7162
s_11_53_09	3.0000	-0.5321	0.8720	2.2674	
s_11_53_10	3.0000	-0.2578	0.8954	4.5262	

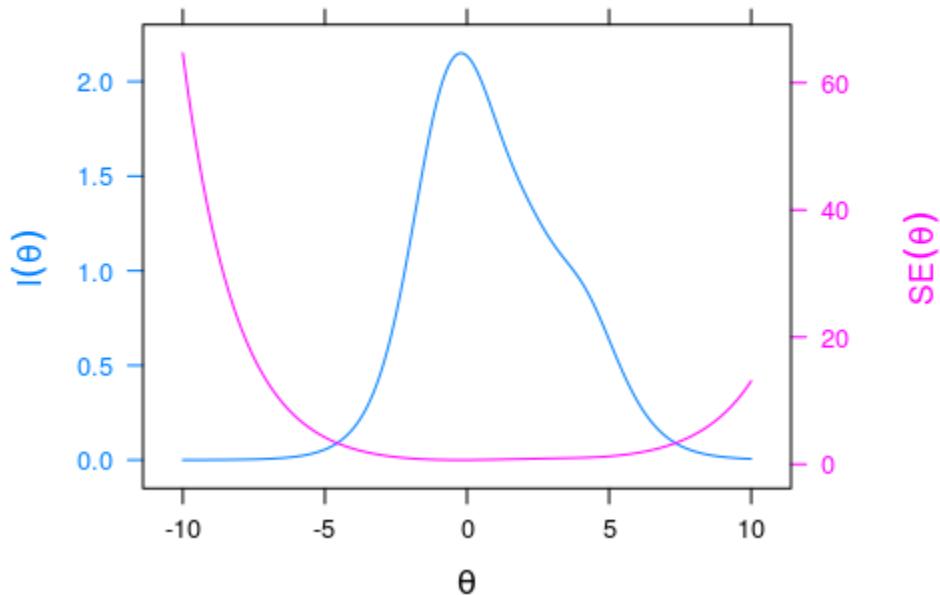
*Table 55: IRT Item Fit (S-X2)*

item	Fit	df	Probability
s_11_53_03	26.775	20	0.142
s_11_53_04	37.190	22	0.023
s_11_53_06	45.419	33	0.073
s_11_53_07	48.440	36	0.081
s_11_53_09	43.188	34	0.134
s_11_53_10	51.081	29	0.007

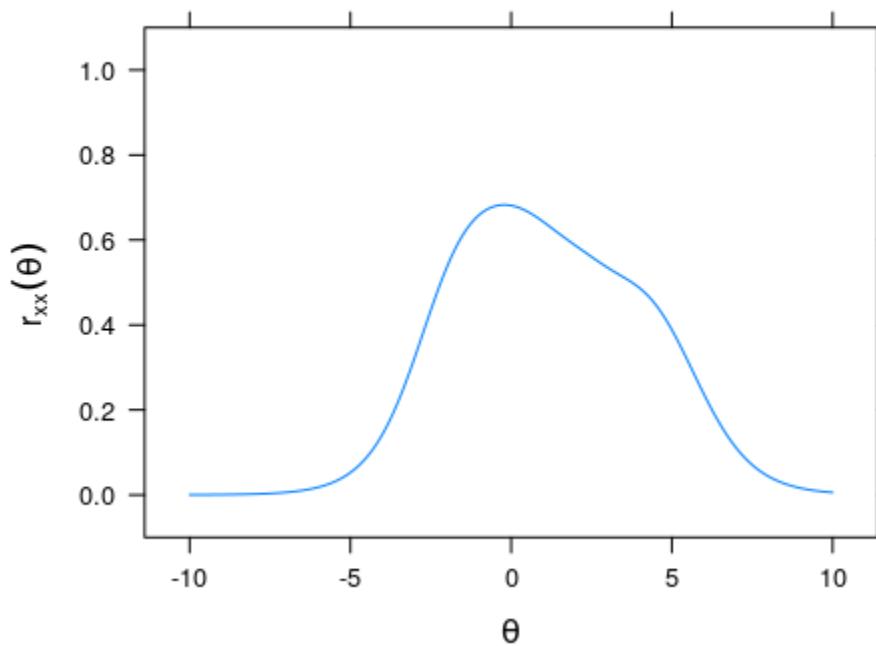
Probability < .01 indicates poor category fit

---

**Test Information (blue) and Standard Error (red)  
by Student Ability**



**Reliability by Student Ability**



## 5.14. Marigolds and Mosquitoes

*Table 56: Item Response Frequencies*

Item	0	1	2	3
s_11_54_02	494	401		
s_11_54_03	509	295	80	11
s_11_54_04	818	61	12	4
s_11_54_07	237	147	140	370
s_11_54_08	459	346	76	14
s_11_54_10	549	200	146	
s_11_54_11	818	69	8	

*Table 57: Item-Total Correlations and Alpha Without*

Variable	Item.Total	Alpha.Without	N
s_11_54_02	0.235	0.479	894
s_11_54_03	0.301	0.446	894
s_11_54_04	0.202	0.492	894
s_11_54_07	0.286	0.496	894
s_11_54_08	0.358	0.421	894
s_11_54_10	0.260	0.463	894
s_11_54_11	0.245	0.488	894

*Table 58: Item Step Locations*

Item	nCats	Step_1	Step_2	Step_3
s_11_54_02	1.0000	0.3523		
s_11_54_03	3.0000	0.3480	3.2096	6.0000
s_11_54_04	3.0000	2.6392	4.4798	6.0000
s_11_54_07	3.0000	-1.5773	-0.3581	0.7985
s_11_54_08	3.0000	-0.0213	2.4392	4.5808
s_11_54_10	2.0000	0.6669	2.8560	
s_11_54_11	2.0000	2.7931	5.4779	

*Table 59: IRT Item Fit (S-X2)*

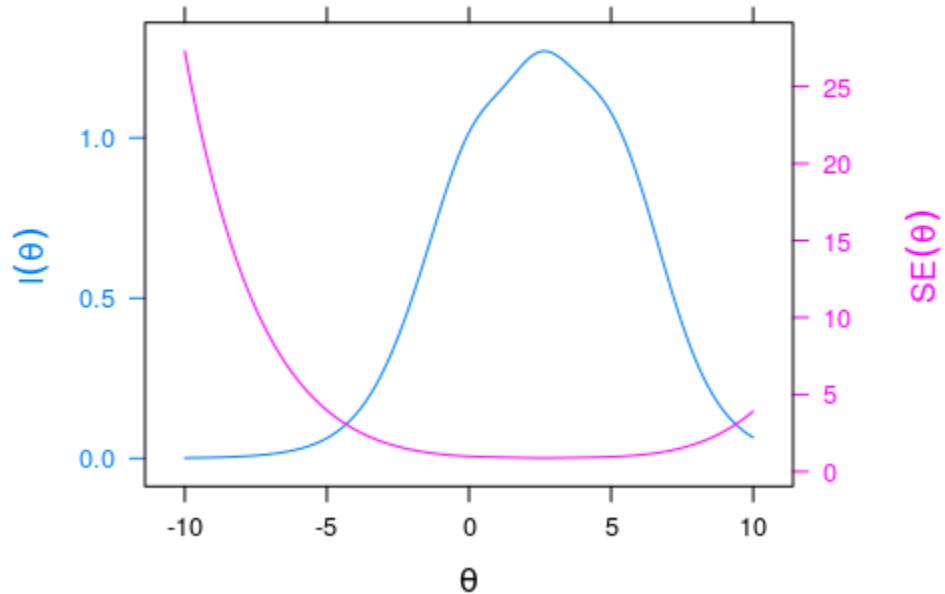
item	Fit	df	Probability
s_11_54_02	8.782	8	0.361
s_11_54_03	26.269	20	0.157
s_11_54_04	21.092	14	0.099
s_11_54_07	48.528	15	0.000

item	Fit	df	Probability
s_11_54_08	29.558	20	0.077
s_11_54_10	17.971	16	0.326
s_11_54_11	16.846	12	0.155

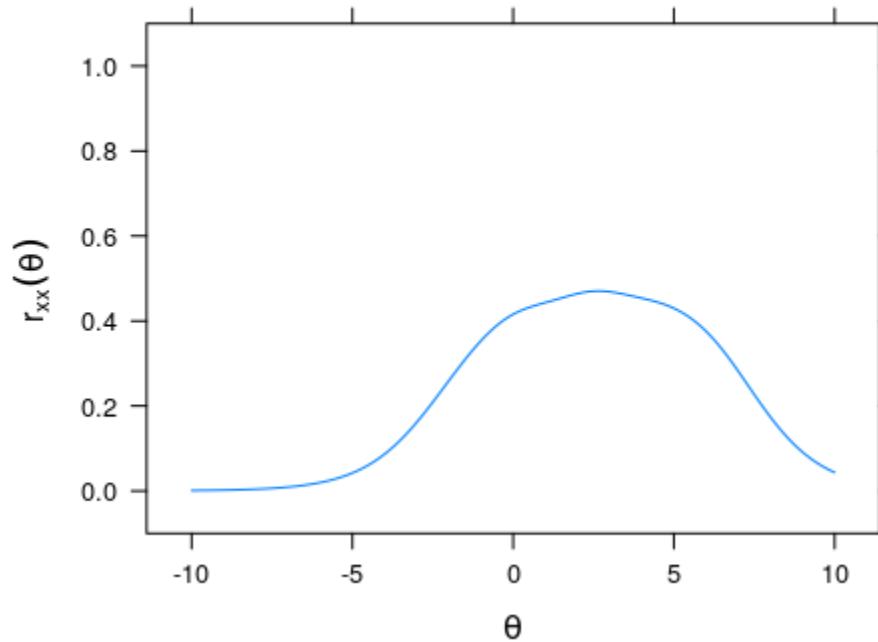
Probability < .01 indicates poor category fit

---

**Test Information (blue) and Standard Error (red)  
by Student Ability**



**Reliability by Student Ability**



## 5.15. Heredity EOU

*Table 60: Item Response Frequencies*

Item	0	1	2	3	4
s_11_55_02	596	78	41	197	
s_11_55_03	107	363	442		
s_11_55_05	590	77	192	53	
s_11_55_06	369	86	133	324	
s_11_55_08	475	317	99	21	
s_11_55_09	353	139	53	365	
s_11_55_10	282	353	67	57	153

*Table 61: Item-Total Correlations and Alpha Without*

Variable	Item.Total	Alpha.Without	N
s_11_55_02	0.313	0.713	910
s_11_55_03	0.495	0.684	910
s_11_55_05	0.476	0.674	910
s_11_55_06	0.486	0.668	910
s_11_55_08	0.441	0.688	910
s_11_55_09	0.415	0.689	910
s_11_55_10	0.500	0.665	910

*Table 62: Item Step Locations*

Item	nCats	Step_1	Step_2	Step_3	Step_4
s_11_55_02	3.0000	0.5756	1.5281	2.4021	
s_11_55_03	2.0000	-1.9424	0.1955		
s_11_55_05	3.0000	0.1329	1.1049	2.5677	
s_11_55_06	3.0000	-0.8419	0.0151	0.9474	
s_11_55_08	3.0000	-0.0232	2.1605	4.2415	
s_11_55_09	3.0000	-0.9054	0.1134	0.9668	
s_11_55_10	4.0000	-0.9639	0.5639	1.2582	2.0515

*Table 63: IRT Item Fit (S-X2)*

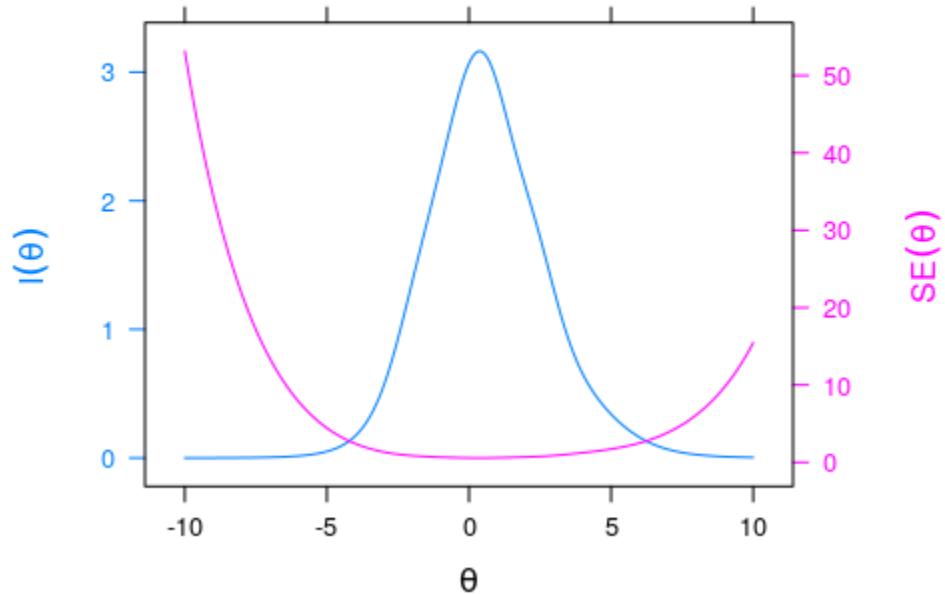
item	Fit	df	Probability
s_11_55_02	51.040	45	0.248
s_11_55_03	13.623	24	0.955
s_11_55_05	47.214	39	0.172

item	Fit	df	Probability
s_11_55_06	49.283	39	0.125
s_11_55_08	28.810	38	0.859
s_11_55_09	45.308	42	0.336
s_11_55_10	55.098	49	0.255

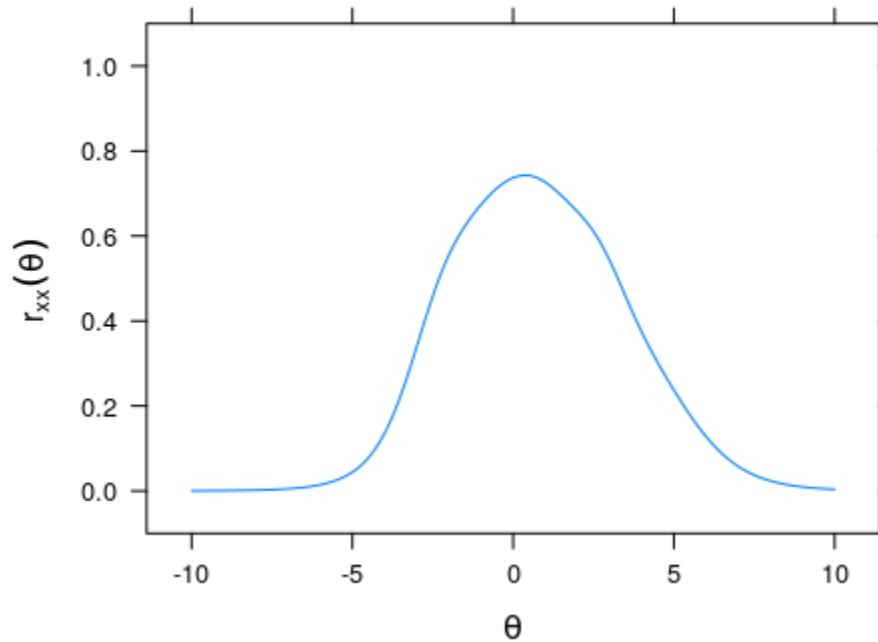
Probability < .01 indicates poor category fit

---

**Test Information (blue) and Standard Error (red)  
by Student Ability**



**Reliability by Student Ability**



## 5.16. Light and waves at the Aquarium

*Table 64: Item Response Frequencies*

Item	0	1	2	3	4
s_15_68_02	427	128	229	130	
s_15_68_03	799	85	29	1	
s_15_68_04	265	550	96	3	
s_15_68_06	90	214	359	251	
s_15_68_08	420	409	85		
s_15_68_09	291	476	99	48	
s_15_68_10	118	114	487	103	92

*Table 65: Item-Total Correlations and Alpha Without*

Variable	Item.Total	Alpha.Without	N
s_15_68_02	0.379	0.587	914
s_15_68_03	0.272	0.618	914
s_15_68_04	0.325	0.602	914
s_15_68_06	0.391	0.576	914
s_15_68_08	0.292	0.609	914
s_15_68_09	0.426	0.567	914
s_15_68_10	0.378	0.584	914

*Table 66: Item Step Locations*

Item	nCats	Step_1	Step_2	Step_3	Step_4
s_15_68_02	3.0000	-0.4770	0.6261	2.0983	
s_15_68_03	3.0000	1.7087	3.3104	5.9349	
s_15_68_04	3.0000	-1.1912	2.7330	6.9155	
s_15_68_06	3.0000	-2.4270	-0.7049	1.1799	
s_15_68_08	2.0000	-0.2514	3.4307		
s_15_68_09	3.0000	-0.9193	1.6681	3.2652	
s_15_68_10	4.0000	-2.5788	-1.1281	1.4299	2.9159

*Table 67: IRT Item Fit (S-X2)*

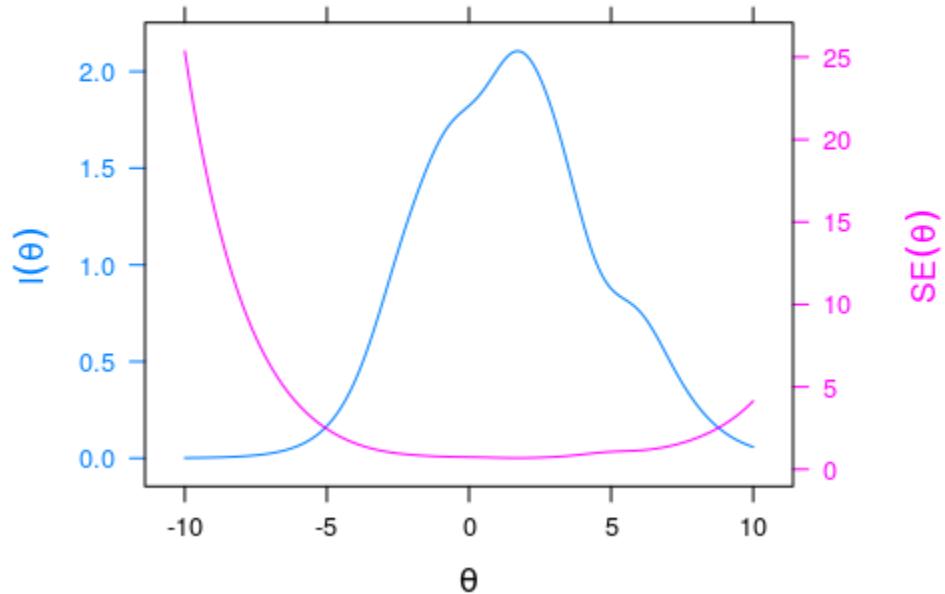
item	Fit	df	Probability
s_15_68_02	20.013	31	0.936
s_15_68_03	21.090	19	0.332
s_15_68_04	24.737	22	0.310

item	Fit	df	Probability
s_15_68_06	25.551	27	0.544
s_15_68_08	23.508	24	0.490
s_15_68_09	34.827	31	0.291
s_15_68_10	48.874	37	0.092

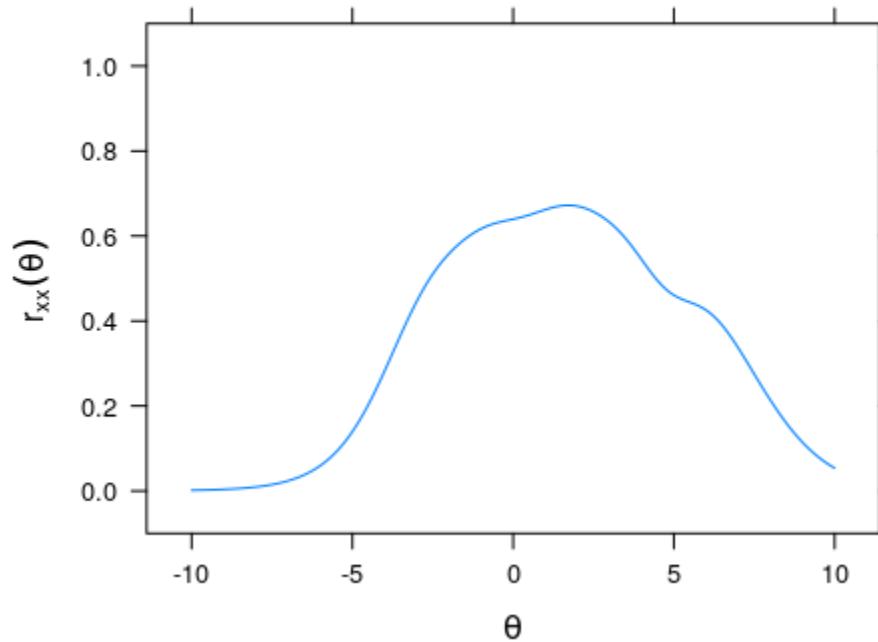
Probability <.01 indicates poor category fit

---

**Test Information (blue) and Standard Error (red)  
by Student Ability**



**Reliability by Student Ability**



## 5.17. Picture Day

*Table 68: Item Response Frequencies*

Item	0	1	2	3	4
s_15_69_02	273	127	296	200	
s_15_69_03	307	478	110	1	
s_15_69_05	104	202	121	469	
s_15_69_07	321	172	332	39	32
s_15_69_09	734	37	61	64	
s_15_69_10	558	133	147	58	

*Table 69: Item-Total Correlations and Alpha Without*

Variable	Item.Total	Alpha.Without	N
s_15_69_02	0.379	0.656	896
s_15_69_03	0.450	0.643	896
s_15_69_05	0.376	0.656	896
s_15_69_07	0.549	0.589	896
s_15_69_09	0.421	0.639	896
s_15_69_10	0.357	0.659	896

*Table 70: Item Step Locations*

Item	nCats	Step_1	Step_2	Step_3	Step_4
s_15_69_02	3.0000	-1.2668	-0.0830	1.6124	
s_15_69_03	3.0000	-0.7029	1.9621	5.7329	
s_15_69_05	3.0000	-2.3765	-0.8704	0.2462	
s_15_69_07	4.0000	-0.8151	0.3922	1.9124	3.0395
s_15_69_09	3.0000	0.7689	1.5518	2.4225	
s_15_69_10	3.0000	0.3876	1.7096	3.5342	

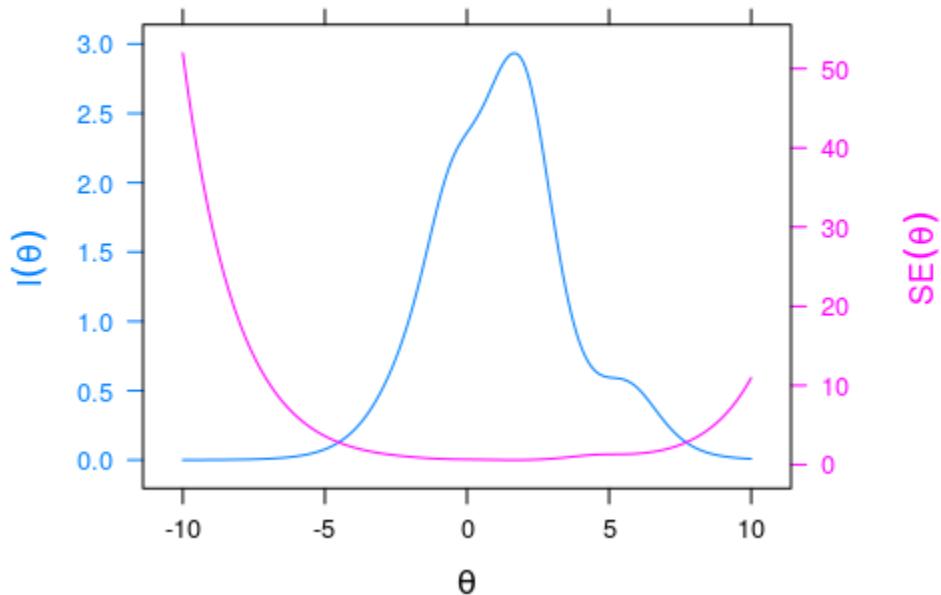
*Table 71: IRT Item Fit (S-X2)*

item	Fit	df	Probability
s_15_69_02	50.166	32	0.021
s_15_69_03	19.261	22	0.629
s_15_69_05	18.412	27	0.890
s_15_69_07	36.938	34	0.335
s_15_69_09	19.014	30	0.940
s_15_69_10	41.775	35	0.200

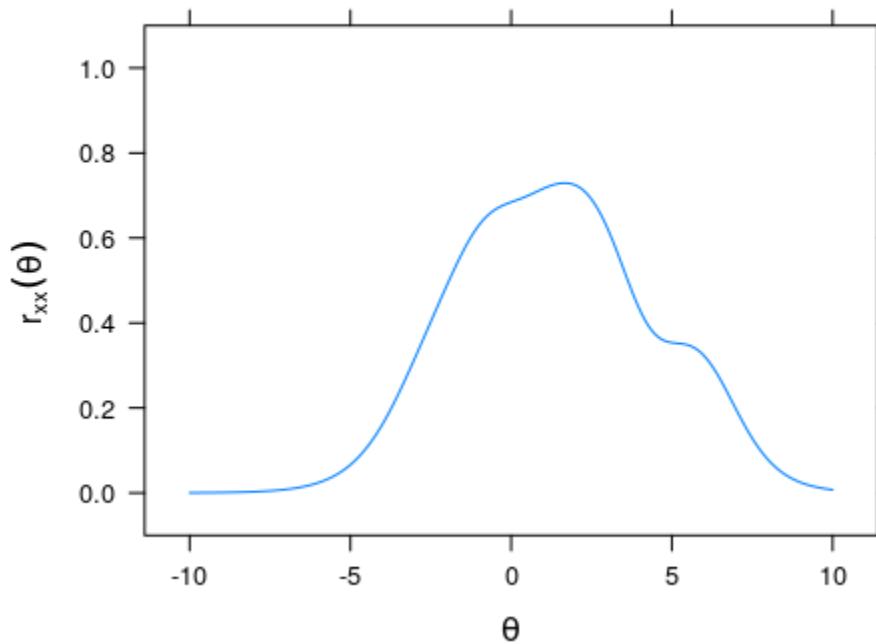
Probability < .01 indicates poor category fit

---

**Test Information (blue) and Standard Error (red)  
by Student Ability**



**Reliability by Student Ability**



## 5.18. Light EOU

*Table 72: Item Response Frequencies*

Item	0	1	2	3	4
s_15_71_02	65	232	270	325	
s_15_71_03	485	282	6	119	
s_15_71_05	271	440	181		
s_15_71_06	178	417	271	26	
s_15_71_08	375	183	280	54	
s_15_71_10	292	145	440	8	7

*Table 73: Item-Total Correlations and Alpha Without*

Variable	Item.Total	Alpha.Without	N
s_15_71_02	0.345	0.644	892
s_15_71_03	0.430	0.612	892
s_15_71_05	0.263	0.664	892
s_15_71_06	0.342	0.643	892
s_15_71_08	0.480	0.592	892
s_15_71_10	0.515	0.579	892

*Table 74: Item Step Locations*

Item	nCats	Step_1	Step_2	Step_3	Step_4
s_15_71_02	3.0000	-3.5200	-1.0331	0.9360	
s_15_71_03	3.0000	-0.0224	1.3050	2.1213	
s_15_71_05	2.0000	-1.7048	2.8954		
s_15_71_06	3.0000	-2.0220	1.0800	4.9385	
s_15_71_08	3.0000	-0.5766	0.7104	2.6411	
s_15_71_10	4.0000	-0.9417	0.2758	2.7330	3.8749

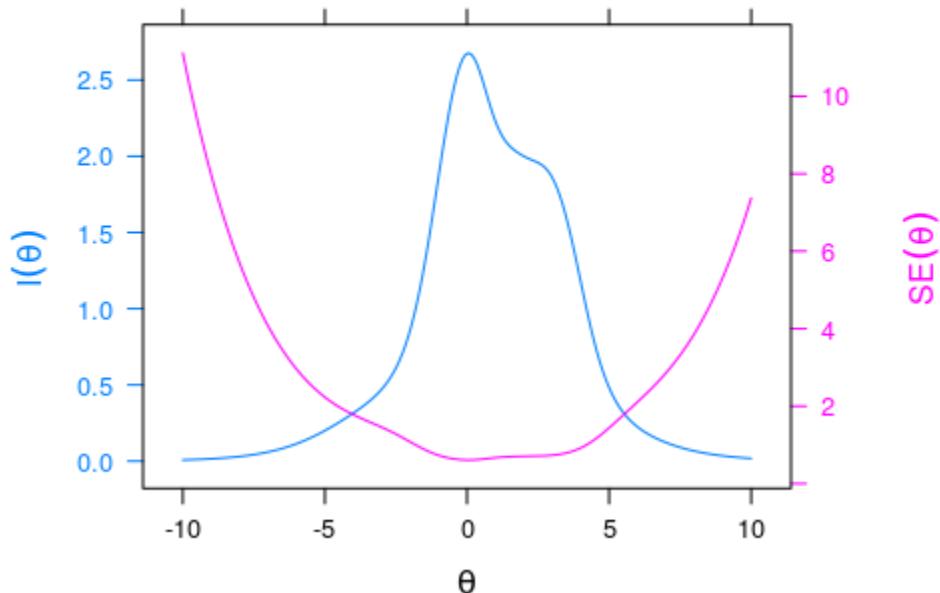
*Table 75: IRT Item Fit (S-X2)*

item	Fit	df	Probability
s_15_71_02	27.570	27	0.433
s_15_71_03	19.952	18	0.336
s_15_71_05	33.097	24	0.102
s_15_71_06	22.878	30	0.820
s_15_71_08	31.137	28	0.311
s_15_71_10	21.493	23	0.551

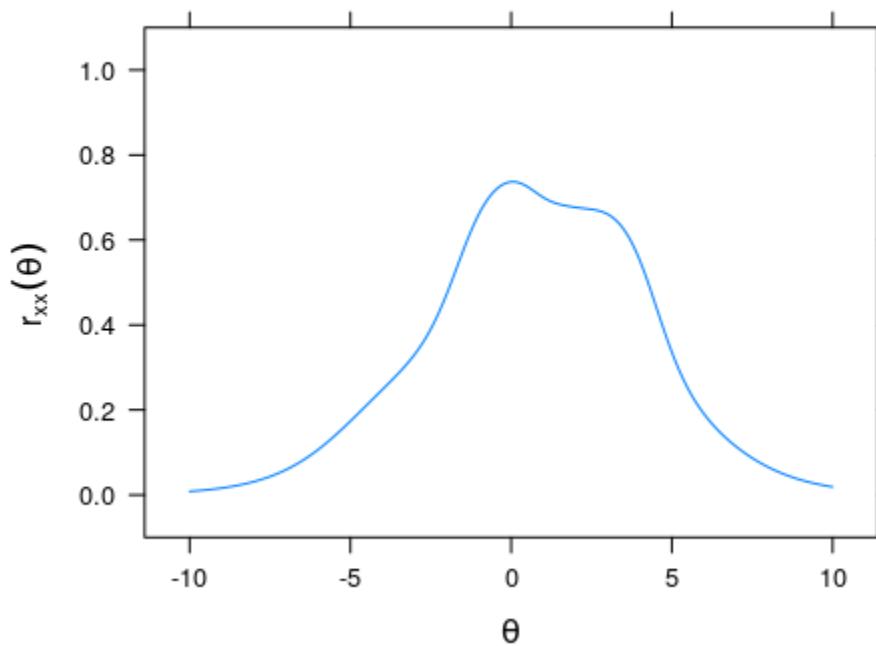
Probability < .01 indicates poor category fit

---

**Test Information (blue) and Standard Error (red)  
by Student Ability**



**Reliability by Student Ability**



## 5.19. School Garden

*Table 76: Item Response Frequencies*

Item	0	1	2	3	4
s_08_41_02	280	222	215	221	
s_08_41_03	447	240	28	213	
s_08_41_04	443	332	130	35	
s_08_41_06	255	507	64	85	29
s_08_41_08	331	334	160	114	
s_08_41_09	284	518	54	84	

*Table 77: Item-Total Correlations and Alpha Without*

Variable	Item.Total	Alpha.Without	N
s_08_41_02	0.372	0.771	925
s_08_41_03	0.510	0.733	925
s_08_41_04	0.454	0.744	925
s_08_41_06	0.585	0.710	925
s_08_41_08	0.569	0.714	925
s_08_41_09	0.619	0.708	925

*Table 78: Item Step Locations*

Item	nCats	Step_1	Step_2	Step_3	Step_4
s_08_41_02	3.0000	-1.3982	0.2026	1.8272	
s_08_41_03	3.0000	-0.3637	0.7512	1.5645	
s_08_41_04	3.0000	-0.3262	1.8060	3.9315	
s_08_41_06	4.0000	-1.0417	1.0092	1.9339	3.2409
s_08_41_08	3.0000	-0.8511	0.7316	2.1311	
s_08_41_09	3.0000	-0.8733	1.2576	2.3785	

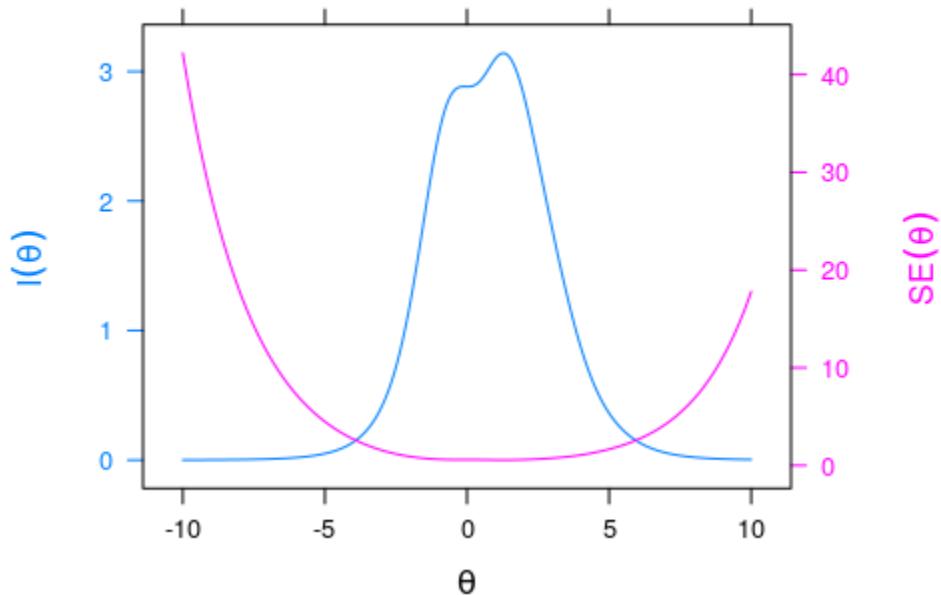
*Table 79: IRT Item Fit (S-X2)*

item	Fit	df	Probability
s_08_41_02	76.588	33	0.000
s_08_41_03	46.332	29	0.022
s_08_41_04	77.114	36	0.000
s_08_41_06	40.174	36	0.290
s_08_41_08	58.643	32	0.003
s_08_41_09	39.313	33	0.208

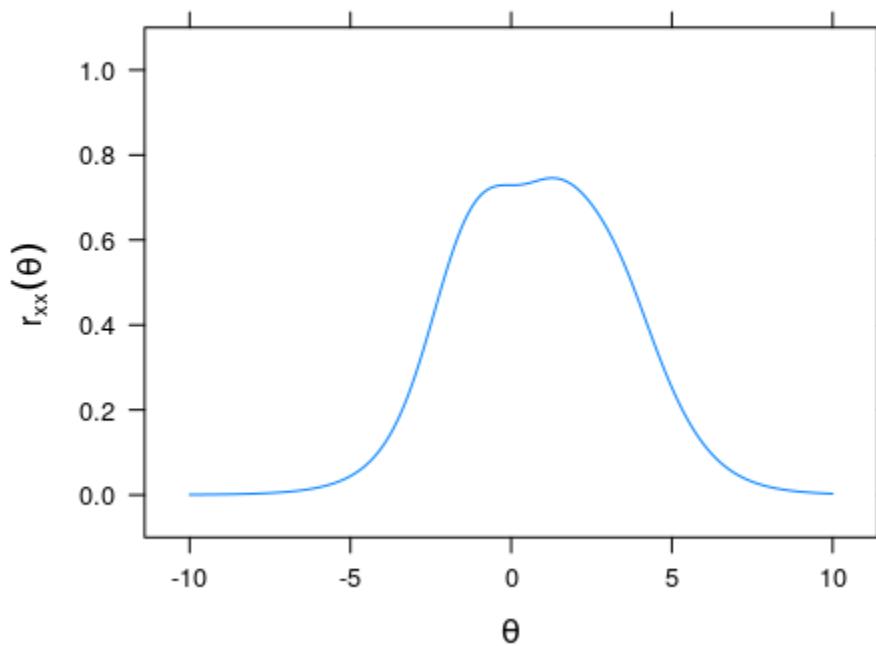
Probability < .01 indicates poor category fit

---

**Test Information (blue) and Standard Error (red)  
by Student Ability**



**Reliability by Student Ability**



## 5.20. Living Things and Food

*Table 80: Item Response Frequencies*

Item	0	1	2	3	4
s_08_42_02	288	355	199		
s_08_42_03	193	100	549		
s_08_42_04	135	389	258	59	
s_08_42_06	561	182	38	61	
s_08_42_07	166	327	291	36	22
s_08_42_08	324	293	192	33	

*Table 81: Item-Total Correlations and Alpha Without*

Variable	Item.Total	Alpha.Without	N
s_08_42_02	0.439	0.634	841
s_08_42_03	0.330	0.667	841
s_08_42_04	0.399	0.645	841
s_08_42_06	0.390	0.648	841
s_08_42_07	0.469	0.620	841
s_08_42_08	0.445	0.629	841

*Table 82: Item Step Locations*

Item	nCats	Step_1	Step_2	Step_3	Step_4
s_08_42_02	2.0000	-0.8759	1.4616		
s_08_42_03	2.0000	-1.8249	-0.5006		
s_08_42_04	3.0000	-2.3936	0.6931	3.6783	
s_08_42_06	3.0000	0.6194	2.2309	3.3791	
s_08_42_07	4.0000	-1.6399	0.3725	2.5231	3.9372
s_08_42_08	3.0000	-0.6668	1.0869	3.1959	

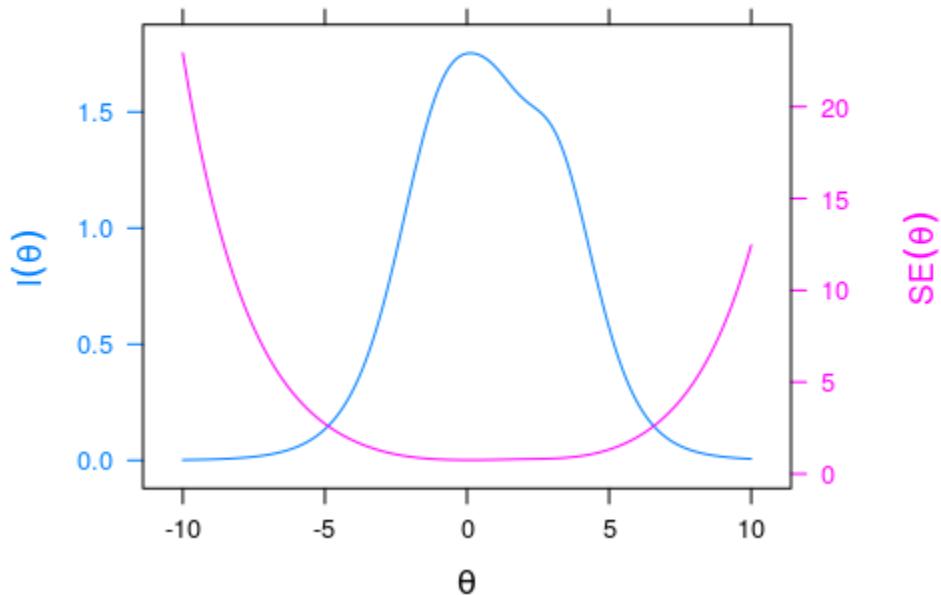
*Table 83: IRT Item Fit (S-X2)*

item	Fit	df	Probability
s_08_42_02	32.813	20	0.035
s_08_42_03	39.147	18	0.003
s_08_42_04	42.619	28	0.038
s_08_42_06	46.219	29	0.022
s_08_42_07	29.048	28	0.410
s_08_42_08	45.009	26	0.012

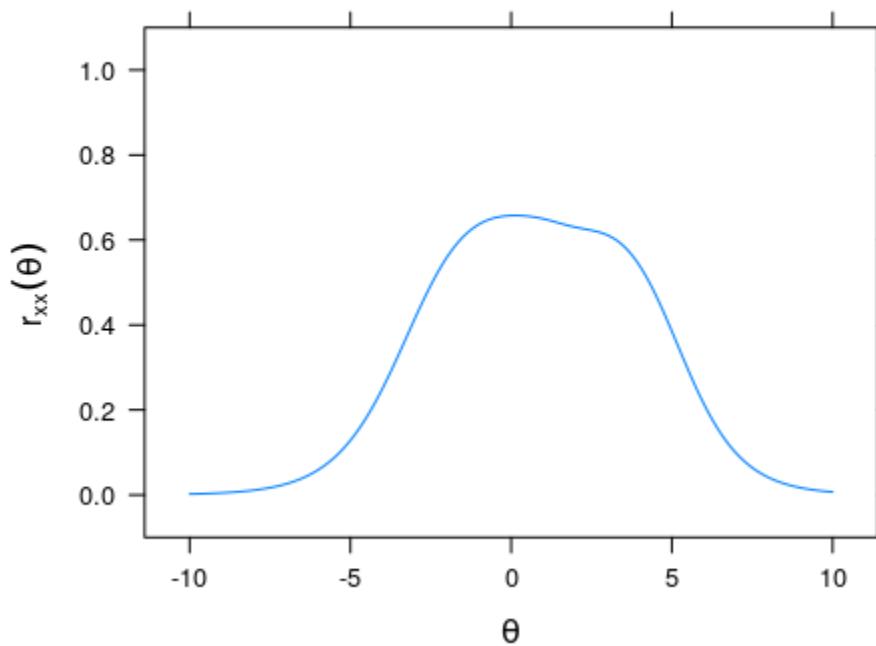
Probability < .01 indicates poor category fit

---

**Test Information (blue) and Standard Error (red)  
by Student Ability**



**Reliability by Student Ability**



## 5.21. Matter and Energy in Living Systems EOU

*Table 84: Item Response Frequencies*

Item	0	1	2	3	4
s_08_43_02	208	124	89	411	
s_08_43_03	142	387	232	70	
s_08_43_04	193	475	97	68	
s_08_43_06	501	221	59	52	
s_08_43_07	153	275	192	155	58
s_08_43_08	268	331	204	28	2

*Table 85: Item-Total Correlations and Alpha Without*

Variable	Item.Total	Alpha.Without	N
s_08_43_02	0.491	0.706	831
s_08_43_03	0.566	0.683	831
s_08_43_04	0.498	0.700	831
s_08_43_06	0.875	0.729	831
s_08_43_07	0.481	0.705	831
s_08_43_08	0.520	0.694	831

*Table 86: Item Step Locations*

Item	nCats	Step_1	Step_2	Step_3	Step_4
s_08_43_02	3.0000	-1.3660	-0.4146	0.4614	
s_08_43_03	3.0000	-1.5816	0.4988	2.3480	
s_08_43_04	3.0000	-1.3189	1.2277	2.6773	
s_08_43_06	3.0000	0.3224	2.0946	3.4442	
s_08_43_07	4.0000	-1.7170	-0.0350	1.2130	2.8262
s_08_43_08	4.0000	-0.8800	0.9831	3.0218	5.1996

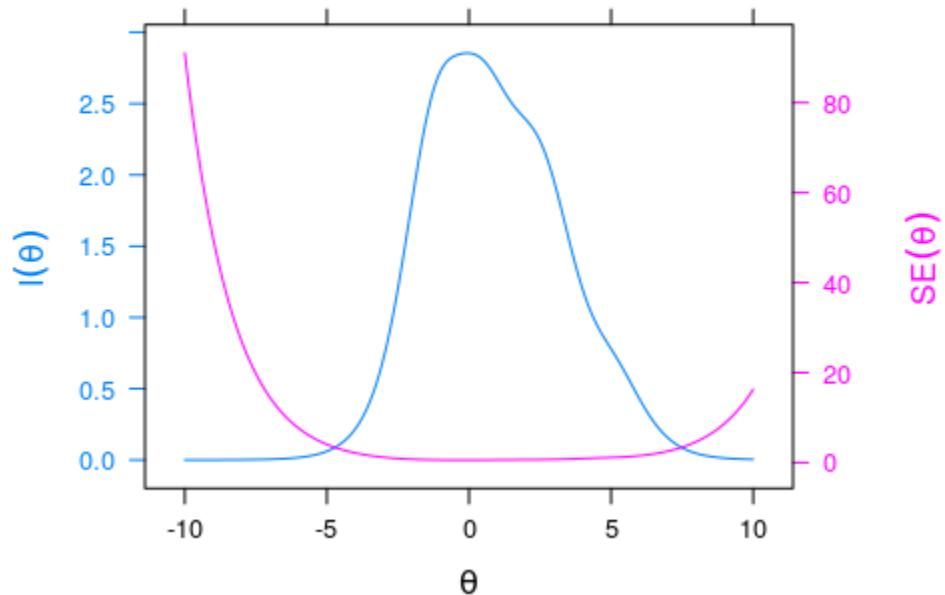
*Table 87: IRT Item Fit (S-X2)*

item	Fit	df	Probability
s_08_43_02	22.175	28	0.773
s_08_43_03	46.026	30	0.031
s_08_43_04	51.657	34	0.027
s_08_43_06	65.982	36	0.002
s_08_43_07	35.929	38	0.566
s_08_43_08	36.473	29	0.160

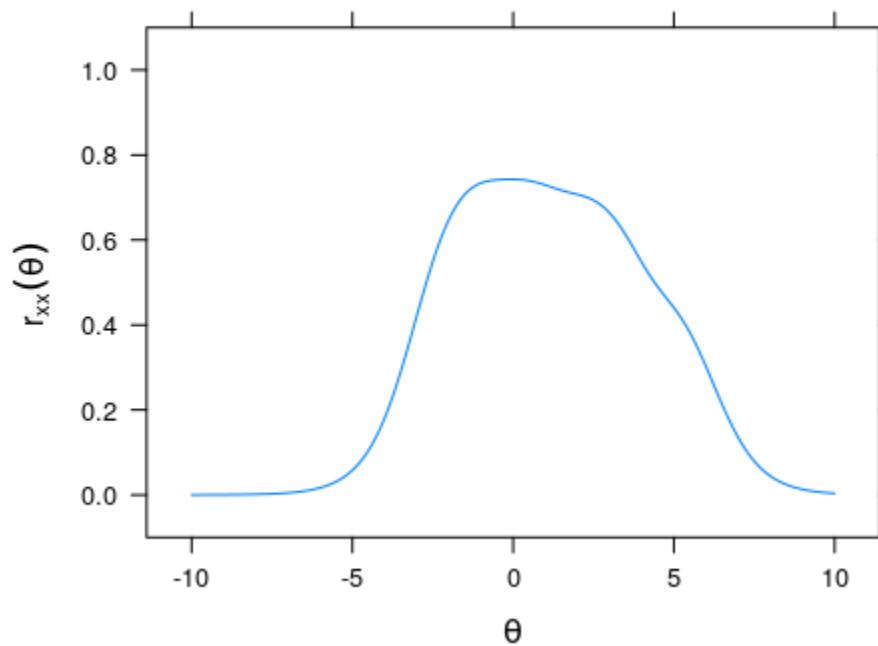
Probability < .01 indicates poor category fit

---

**Test Information (blue) and Standard Error (red)  
by Student Ability**



**Reliability by Student Ability**



## 5.22. Space Station

*Table 88: Item Response Frequencies*

Item	0	1	2	3
s_04_14_02	542	737	370	
s_04_14_03	361	717	572	
s_04_14_05	769	262	619	
s_04_14_06	337	898	136	277
s_04_14_08	1140	196	283	
s_04_14_09	1233	417		
s_04_14_10	670	631	344	
s_04_14_11	432	873	340	
s_04_14_12	1160	380	110	

*Table 89: Item-Total Correlations and Alpha Without*

Variable	Item.Total	Alpha.Without	N
s_04_14_02	0.351	0.704	1607
s_04_14_03	0.388	0.698	1607
s_04_14_05	0.490	0.677	1607
s_04_14_06	0.447	0.688	1607
s_04_14_08	0.458	0.685	1607
s_04_14_09	0.301	0.713	1607
s_04_14_10	0.494	0.678	1607
s_04_14_11	0.467	0.685	1607
s_04_14_12	0.160	0.731	1607

*Table 90: Item Step Locations*

Item	nCats	Step_1	Step_2	Step_3
s_04_14_02	2.0000	-0.9395	1.8957	
s_04_14_03	2.0000	-1.6466	1.0454	
s_04_14_05	2.0000	-0.3637	0.9940	
s_04_14_06	3.0000	-1.7268	1.2543	2.5024
s_04_14_08	2.0000	0.7048	2.0211	
s_04_14_09	1.0000	1.8518		
s_04_14_10	2.0000	-0.4813	1.8822	
s_04_14_11	2.0000	-1.1225	1.7137	
s_04_14_12	2.0000	2.2661	6.0000	

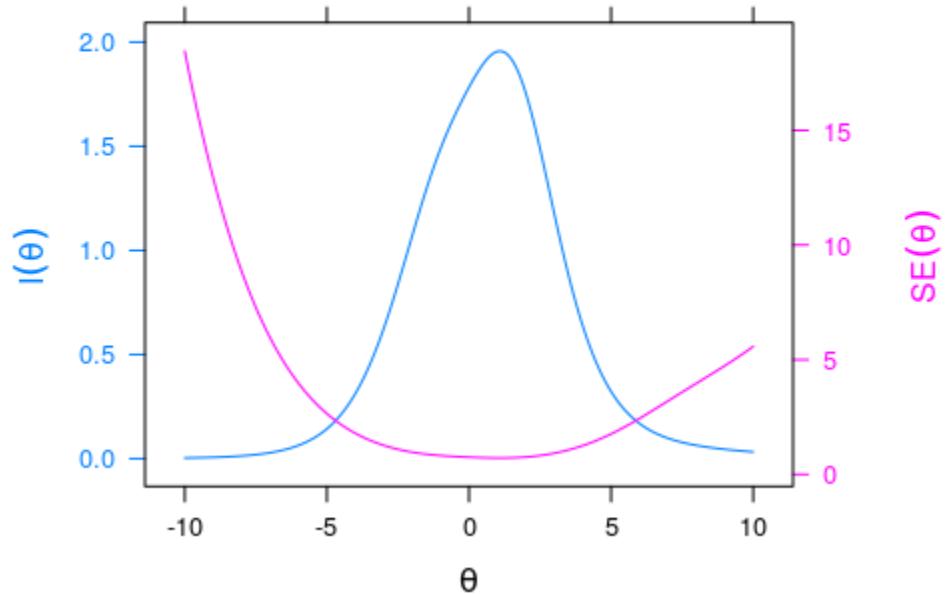
*Table 91: IRT Item Fit (S-X2)*

item	Fit	df	Probability
s_04_14_02	49.820	26	0.003
s_04_14_03	40.587	24	0.018
s_04_14_05	48.777	25	0.003
s_04_14_06	64.838	35	0.002
s_04_14_08	31.853	25	0.162
s_04_14_09	53.808	14	0.000
s_04_14_10	43.439	25	0.013
s_04_14_11	38.112	24	0.034
s_04_14_12	86.146	27	0.000

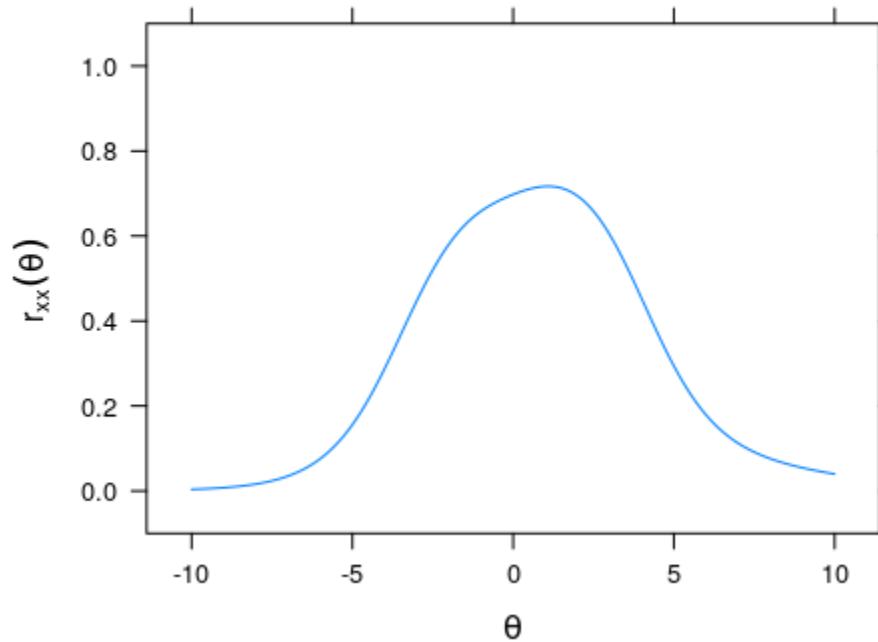
Probability < .01 indicates poor category fit

---

**Test Information (blue) and Standard Error (red)  
by Student Ability**



**Reliability by Student Ability**



## 5.23. Apples

*Table 92: Item Response Frequencies*

Item	0	1	2	3
s_04_15_02	1025	889		
s_04_15_03	418	471	799	226
s_04_15_04	399	972	540	
s_04_15_05	1004	907		
s_04_15_06	948	965		
s_04_15_08	551	1104	29	230
s_04_15_09	537	785	450	141
s_04_15_10	1130	320	371	93

*Table 93: Item-Total Correlations and Alpha Without*

Variable	Item.Total	Alpha.Without	N
s_04_15_02	0.217	0.684	1907
s_04_15_03	0.504	0.619	1907
s_04_15_04	0.417	0.646	1907
s_04_15_05	0.278	0.675	1907
s_04_15_06	0.149	0.694	1907
s_04_15_08	0.498	0.621	1907
s_04_15_09	0.456	0.634	1907
s_04_15_10	0.434	0.641	1907

*Table 94: Item Step Locations*

Item	nCats	Step_1	Step_2	Step_3
s_04_15_02	1.0000	0.2977		
s_04_15_03	3.0000	-1.5089	-0.0276	2.0310
s_04_15_04	2.0000	-1.5115	1.1363	
s_04_15_05	1.0000	0.1744		
s_04_15_06	1.0000	-0.0630		
s_04_15_08	3.0000	-0.9937	1.4532	2.4182
s_04_15_09	3.0000	-1.2426	0.9382	3.0048
s_04_15_10	3.0000	0.1115	1.3618	3.0902

*Table 95: IRT Item Fit (S-X2)*

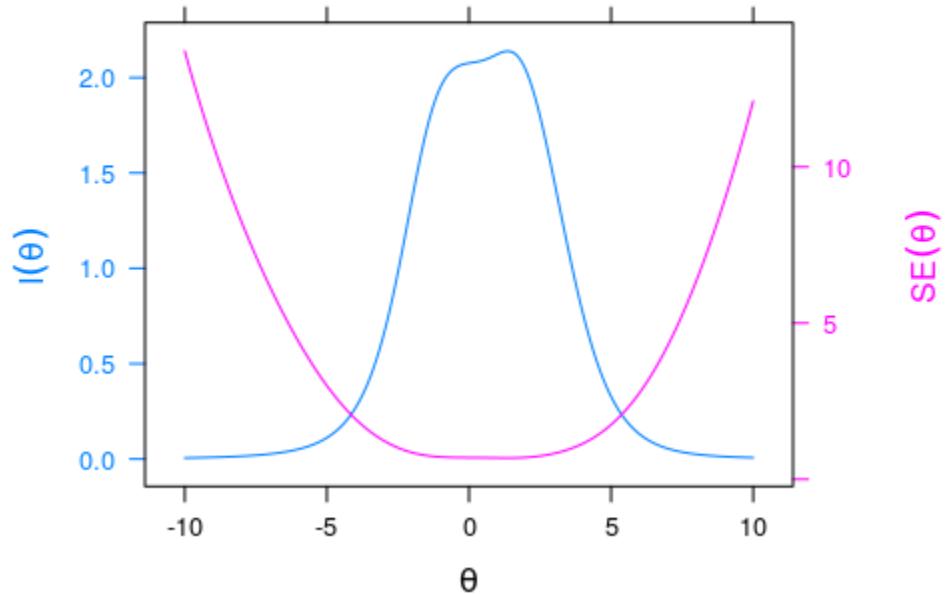
item	Fit	df	Probability
s_04_15_02	24.011	13	0.031

item	Fit	df	Probability
s_04_15_03	30.388	29	0.395
s_04_15_04	87.004	23	0.000
s_04_15_05	23.488	13	0.036
s_04_15_06	13.676	13	0.397
s_04_15_08	40.251	29	0.080
s_04_15_09	44.228	31	0.058
s_04_15_10	87.908	30	0.000

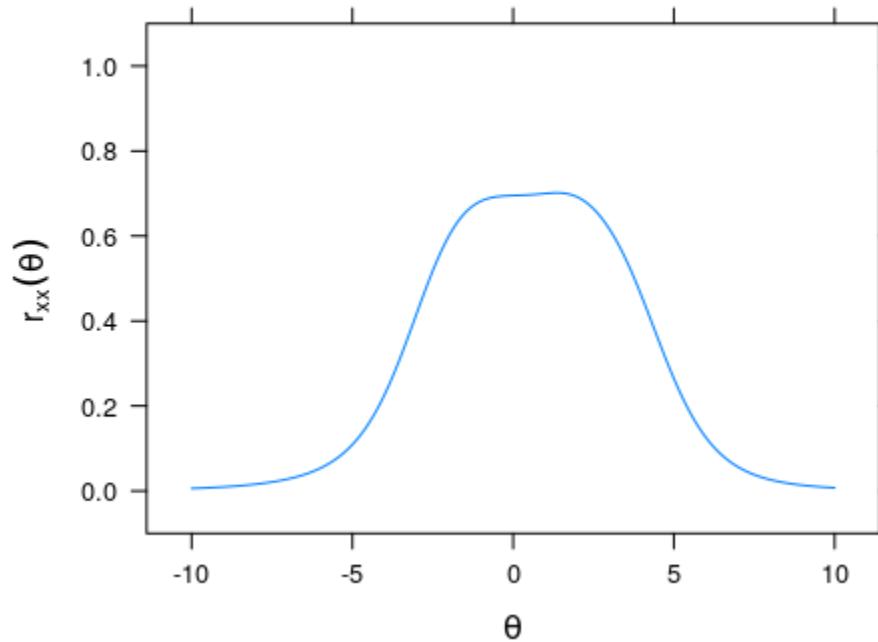
Probability < .01 indicates poor category fit

---

**Test Information (blue) and Standard Error (red)  
by Student Ability**



**Reliability by Student Ability**



## 5.24. Motion and Forces EOU

*Table 96: Item Response Frequencies*

Item	0	1	2	3
s_04_17_02	966	678		
s_04_17_03	1003	639		
s_04_17_04	299	479	867	
s_04_17_05	807	515	322	
s_04_17_06	983	305	357	
s_04_17_07	924	316	304	101
s_04_17_08	919	676	32	17
s_04_17_10	182	884	576	
s_04_17_11	1144	90	411	

*Table 97: Item-Total Correlations and Alpha Without*

Variable	Item.Total	Alpha.Without	N
s_04_17_02	0.363	0.706	1636
s_04_17_03	0.390	0.703	1636
s_04_17_04	0.379	0.701	1636
s_04_17_05	0.574	0.663	1636
s_04_17_06	0.516	0.674	1636
s_04_17_07	0.454	0.689	1636
s_04_17_08	0.352	0.706	1636
s_04_17_10	0.291	0.715	1636
s_04_17_11	0.314	0.717	1636

*Table 98: Item Step Locations*

Item	nCats	Step_1	Step_2	Step_3
s_04_17_02	1.0000	0.5346		
s_04_17_03	1.0000	0.6297		
s_04_17_04	2.0000	-1.8011	0.1339	
s_04_17_05	2.0000	-0.1233	1.6411	
s_04_17_06	2.0000	0.2617	1.6772	
s_04_17_07	3.0000	0.1485	1.5376	3.3595
s_04_17_08	3.0000	0.4476	4.8568	6.0000
s_04_17_10	2.0000	-3.4055	1.1717	
s_04_17_11	2.0000	0.9446	2.0994	

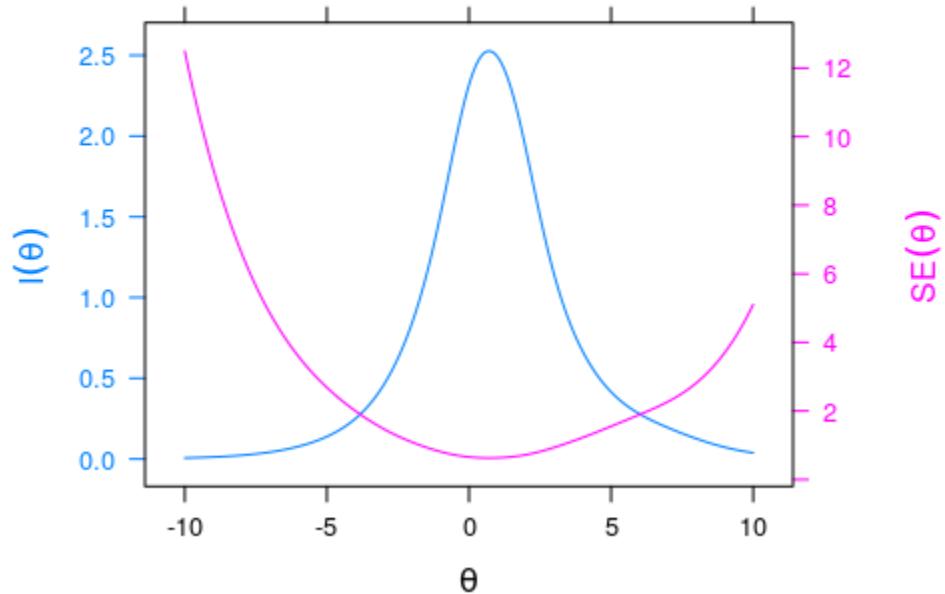
*Table 99: IRT Item Fit (S-X2)*

item	Fit	df	Probability
s_04_17_02	22.485	13	0.048
s_04_17_03	9.316	13	0.749
s_04_17_04	49.034	22	0.001
s_04_17_05	44.717	23	0.004
s_04_17_06	63.570	23	0.000
s_04_17_07	53.737	32	0.009
s_04_17_08	50.607	31	0.015
s_04_17_10	22.144	24	0.571
s_04_17_11	33.333	25	0.123

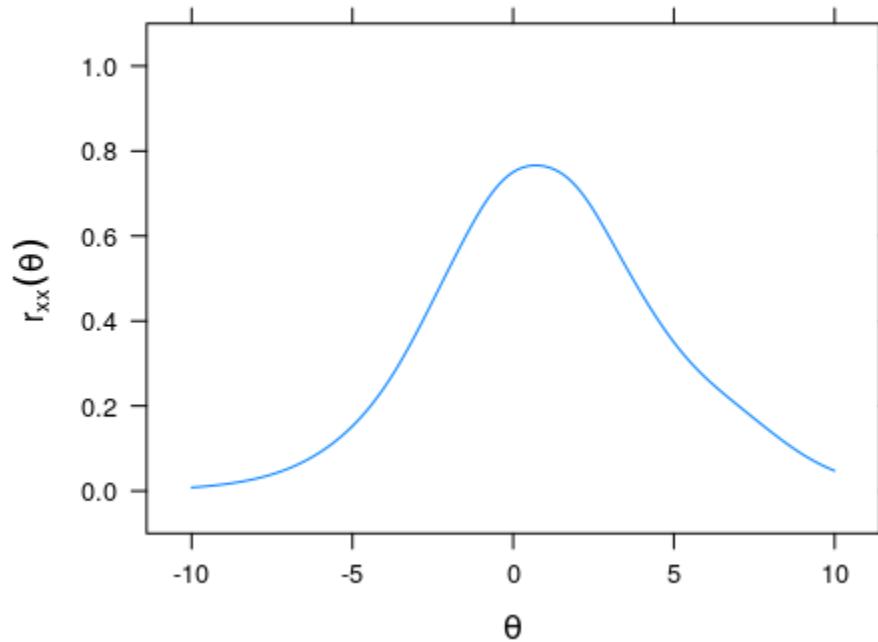
Probability < .01 indicates poor category fit

---

**Test Information (blue) and Standard Error (red)  
by Student Ability**



**Reliability by Student Ability**



## 5.25. Matter in Natural Resources Rescored

*Table 100: Item Response Frequencies*

Item	0	1	2	3	4
s_10_48_02	236	418	261	115	
s_10_48_03	246	446	272	39	
s_10_48_04	147	188	348	320	
s_10_48_05	99	444	226	262	
s_10_48_07	308	563	149		
s_10_48_08	150	268	365	180	40
s_10_48_10	169	557	305		
s_10_48_11	637	229	165		

*Table 101: Item-Total Correlations and Alpha Without*

Variable	Item.Total	Alpha.Without	N
s_10_48_02	0.347	0.637	993
s_10_48_03	0.398	0.625	993
s_10_48_04	0.226	0.674	993
s_10_48_05	0.440	0.611	993
s_10_48_07	0.409	0.628	993
s_10_48_08	0.378	0.630	993
s_10_48_10	0.396	0.630	993
s_10_48_11	0.339	0.639	993

*Table 102: Item Step Locations*

Item	nCats	Step_1	Step_2	Step_3	Step_4
s_10_48_02	3.0000	-1.6820	0.7332	2.8866	
s_10_48_03	3.0000	-1.4838	1.0489	3.9209	
s_10_48_04	3.0000	-3.7433	-1.3703	1.6513	
s_10_48_05	3.0000	-2.5762	-0.0238	1.5057	
s_10_48_07	2.0000	-1.0225	2.1379		
s_10_48_08	4.0000	-2.9688	-0.5307	2.1559	5.2564
s_10_48_10	2.0000	-1.8875	1.1167		
s_10_48_11	2.0000	0.2677	1.8412		

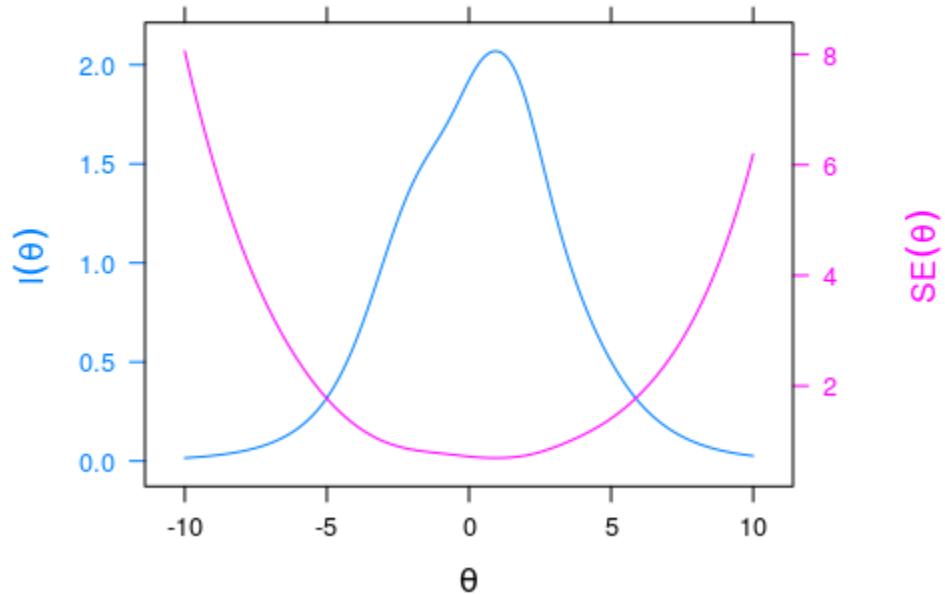
*Table 103: IRT Item Fit (S-X2)*

item	Fit	df	Probability
s_10_48_02	56.181	39	0.037
s_10_48_03	34.600	39	0.671
s_10_48_04	78.210	41	0.000
s_10_48_05	50.186	36	0.058
s_10_48_07	25.422	28	0.605
s_10_48_08	54.805	48	0.232
s_10_48_10	39.226	26	0.046
s_10_48_11	56.228	26	0.001

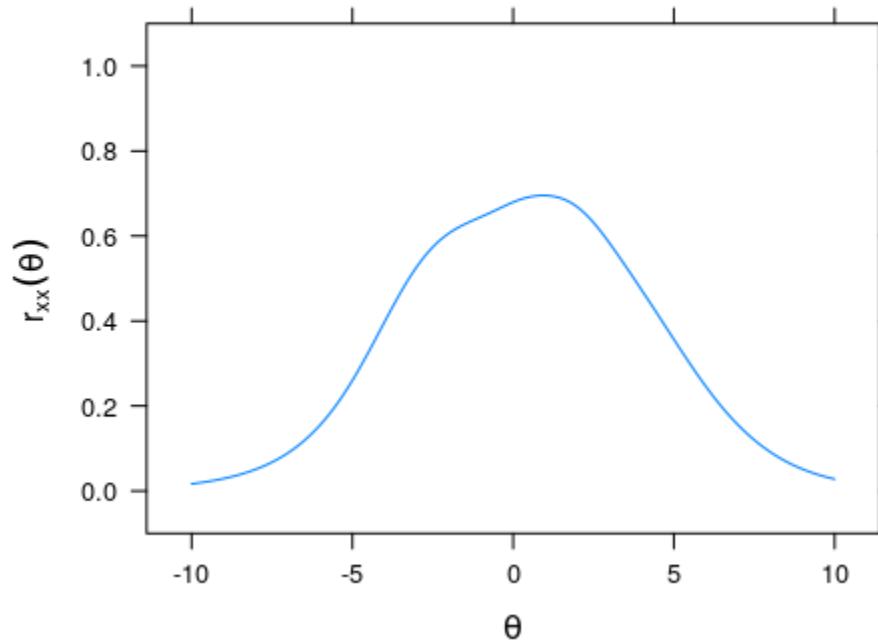
Probability < .01 indicates poor category fit

---

**Test Information (blue) and Standard Error (red)  
by Student Ability**



**Reliability by Student Ability**



## 5.26. Matter in the Kitchen Rescored

*Table 104: Item Response Frequencies*

Item	0	1	2	3	4
s_10_49_02	364	127	354	161	
s_10_49_04	42	125	250	260	310
s_10_49_05	171	257	302	284	
s_10_49_07	285	355	347		
s_10_49_08	579	206	202		
s_10_49_10	135	239	640		
s_10_49_11	114	298	384	218	

*Table 105: Item-Total Correlations and Alpha Without*

Variable	Item.Total	Alpha.Without	N
s_10_49_02	0.433	0.613	979
s_10_49_04	0.483	0.595	979
s_10_49_05	0.352	0.639	979
s_10_49_07	0.393	0.629	979
s_10_49_08	0.410	0.624	979
s_10_49_10	0.187	0.674	979
s_10_49_11	0.375	0.631	979

*Table 106: Item Step Locations*

Item	nCats	Step_1	Step_2	Step_3	Step_4
s_10_49_02	3.0000	-0.9493	0.2157	2.0914	
s_10_49_04	4.0000	-2.8680	-1.4575	-0.2936	0.9967
s_10_49_05	3.0000	-2.1693	-0.3841	1.4081	
s_10_49_07	2.0000	-1.2367	0.8403		
s_10_49_08	2.0000	0.1006	1.5996		
s_10_49_10	2.0000	-4.7101	-1.2570		
s_10_49_11	3.0000	-2.4545	-0.4089	1.6692	

*Table 107: IRT Item Fit (S-X2)*

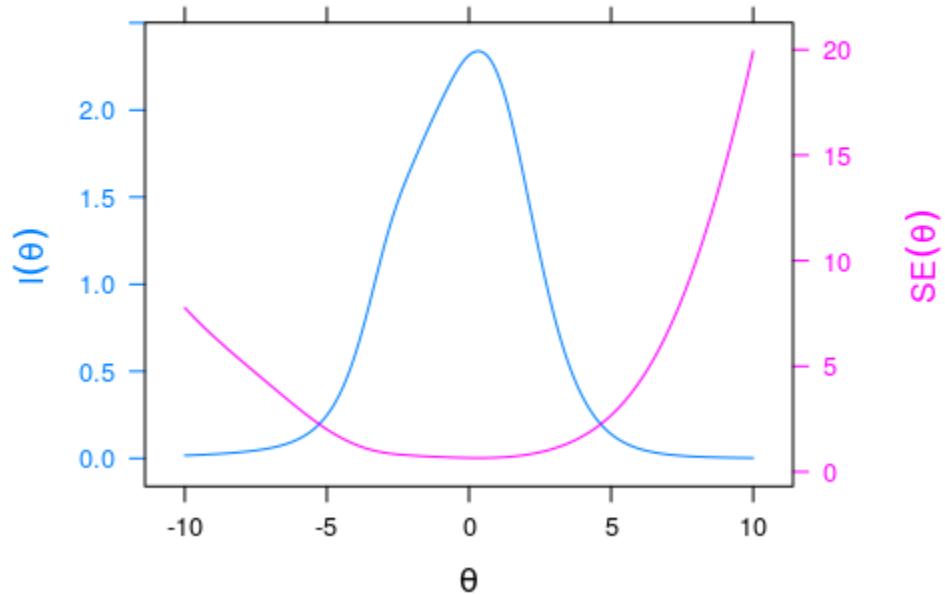
item	Fit	df	Probability
s_10_49_02	37.207	35	0.368
s_10_49_04	22.003	37	0.976
s_10_49_05	30.050	37	0.784

item	Fit	df	Probability
s_10_49_07	26.079	27	0.514
s_10_49_08	15.275	24	0.912
s_10_49_10	69.714	29	0.000
s_10_49_11	30.956	35	0.664

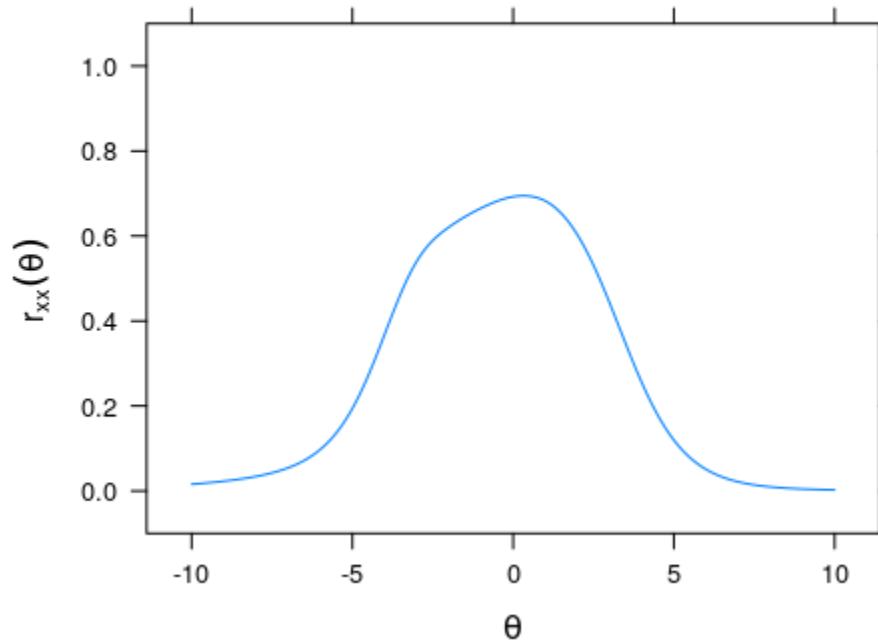
Probability < .01 indicates poor category fit

---

**Test Information (blue) and Standard Error (red)  
by Student Ability**



**Reliability by Student Ability**



## 5.27. Particle Nature of Matter EOU Rescored

*Table 108: Item Response Frequencies*

Item	0	1	2	3	4
s_10_51_03	101	316	397	154	
s_10_51_04	297	468	140	36	
s_10_51_06	211	145	217	278	117
s_10_51_07	167	260	192	349	
s_10_51_09	203	517	221		
s_10_51_10	383	155	301	102	

*Table 109: Item-Total Correlations and Alpha Without*

Variable	Item.Total	Alpha.Without	N
s_10_51_03	0.382	0.626	941
s_10_51_04	0.431	0.615	941
s_10_51_06	0.393	0.634	941
s_10_51_07	0.385	0.625	941
s_10_51_09	0.408	0.627	941
s_10_51_10	0.447	0.600	941

*Table 110: Item Step Locations*

Item	nCats	Step_1	Step_2	Step_3	Step_4
s_10_51_03	3.0000	-2.8058	-0.3216	2.2792	
s_10_51_04	3.0000	-0.8329	1.2795	2.9732	
s_10_51_06	4.0000	-1.7702	-0.5655	0.6450	2.5272
s_10_51_07	3.0000	-1.9335	-0.3381	0.9656	
s_10_51_09	2.0000	-1.5469	1.4067		
s_10_51_10	3.0000	-0.7454	0.5034	2.4487	

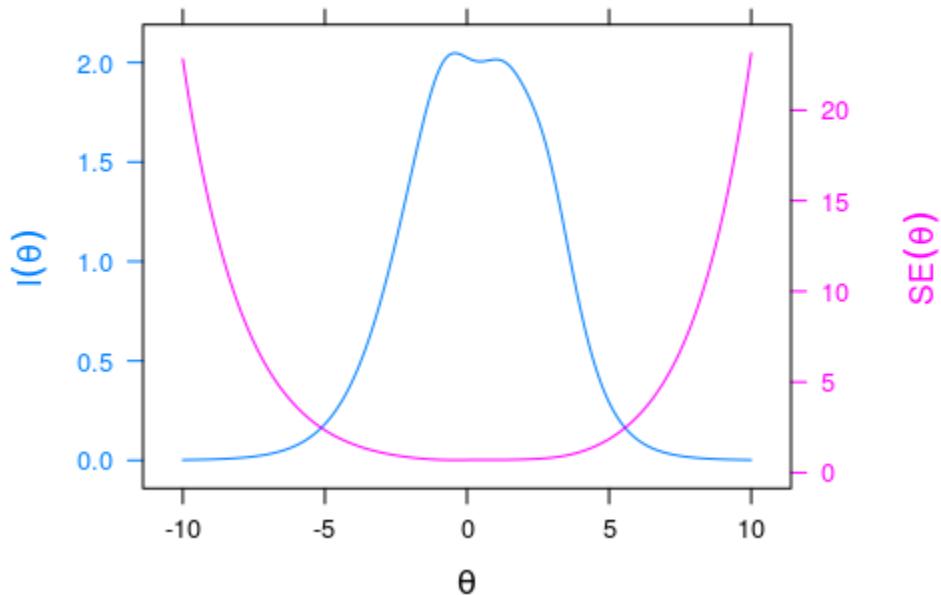
*Table 111: IRT Item Fit (S-X2)*

item	Fit	df	Probability
s_10_51_03	36.784	33	0.298
s_10_51_04	30.689	28	0.331
s_10_51_06	37.936	37	0.427
s_10_51_07	66.893	32	0.000
s_10_51_09	28.341	24	0.246
s_10_51_10	36.875	33	0.294

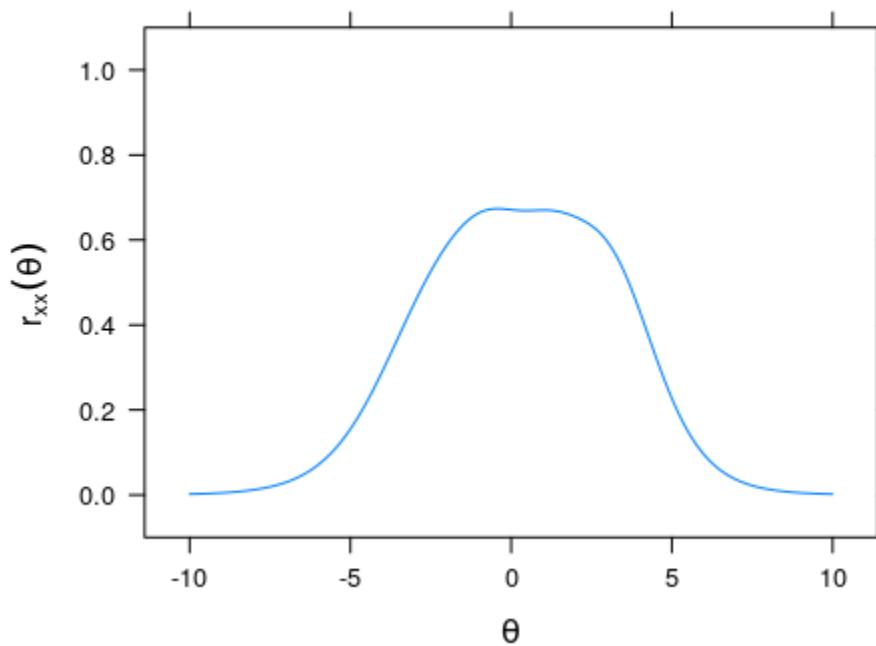
Probability < .01 indicates poor category fit

---

**Test Information (blue) and Standard Error (red)  
by Student Ability**



**Reliability by Student Ability**



## 5.28. Changing Continents

*Table 112: Item Response Frequencies*

Item	0	1	2	3
s_13_60_02	149	138	680	94
s_13_60_03	449	159	453	
s_13_60_04	284	143	477	157
s_13_60_06	520	226	235	80
s_13_60_07	367	145	313	236
s_13_60_08	449	282	206	124
s_13_60_10	491	150	302	118

*Table 113: Item-Total Correlations and Alpha Without*

Variable	Item.Total	Alpha.Without	N
s_13_60_02	0.357	0.700	1061
s_13_60_03	0.477	0.673	1061
s_13_60_04	0.467	0.673	1061
s_13_60_06	0.505	0.664	1061
s_13_60_07	0.469	0.673	1061
s_13_60_08	0.269	0.722	1061
s_13_60_10	0.459	0.675	1061

*Table 114: Item Step Locations*

Item	nCats	Step_1	Step_2	Step_3
s_13_60_02	3.0000	-2.7486	-1.1618	3.2414
s_13_60_03	2.0000	-0.6413	0.6574	
s_13_60_04	3.0000	-1.5746	-0.2584	2.2825
s_13_60_06	3.0000	-0.2208	1.4837	3.8975
s_13_60_07	3.0000	-1.1740	0.0552	1.8212
s_13_60_08	3.0000	-0.6870	1.4070	3.6386
s_13_60_10	3.0000	-0.4853	0.6673	2.3713

*Table 115: IRT Item Fit (S-X2)*

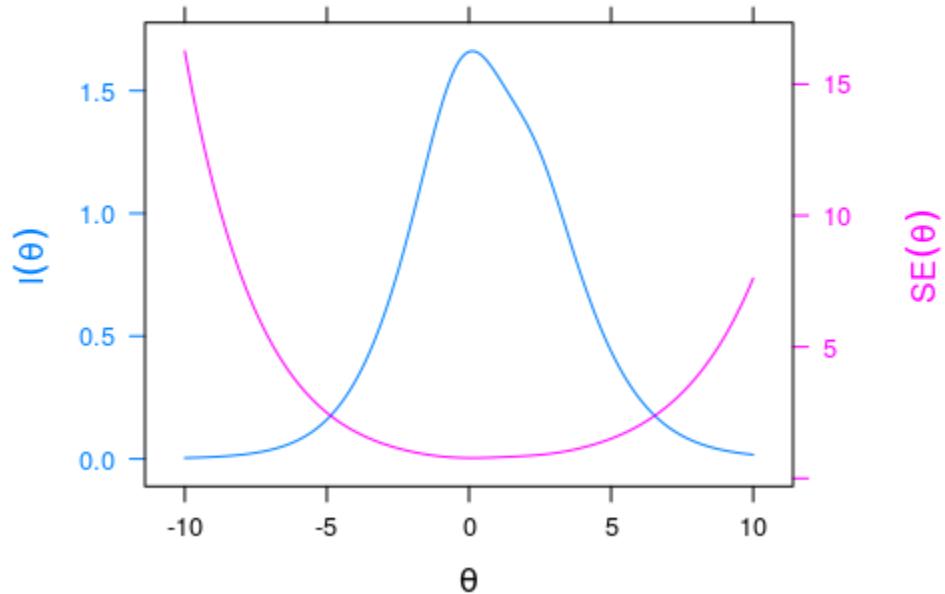
item	Fit	df	Probability
s_13_60_02	48.420	35	0.065
s_13_60_03	28.519	27	0.385
s_13_60_04	43.785	36	0.175
s_13_60_06	66.947	40	0.005

item	Fit	df	Probability
s_13_60_07	55.098	40	0.056
s_13_60_08	50.681	41	0.143
s_13_60_10	47.512	37	0.115

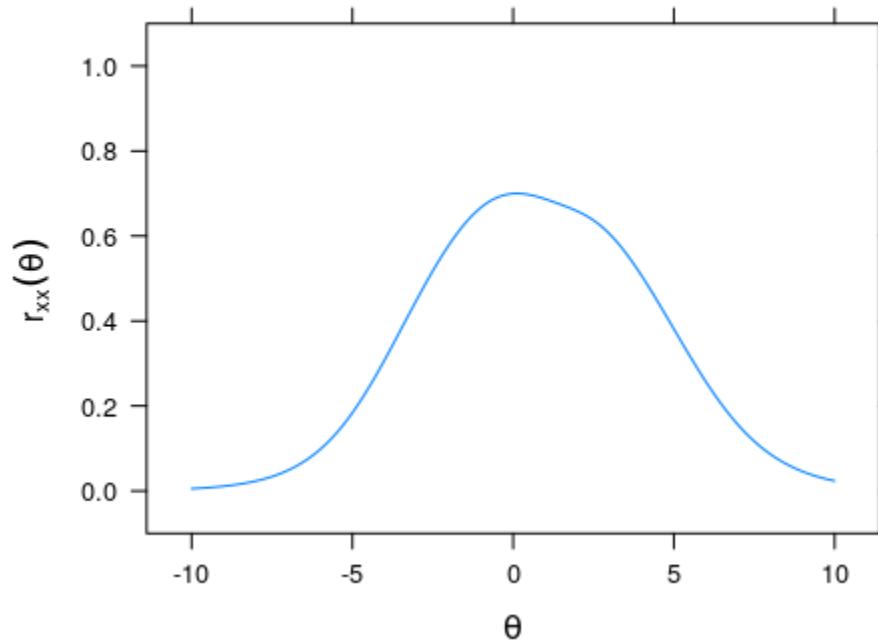
Probability < .01 indicates poor category fit

---

**Test Information (blue) and Standard Error (red)  
by Student Ability**



**Reliability by Student Ability**



## 5.29. Dynamic Earth

*Table 116: Item Response Frequencies*

Item	0	1	2	3
s_13_61_02	131	100	593	202
s_13_61_03	346	309	267	103
s_13_61_05	389	375	191	71
s_13_61_06	135	311	241	339
s_13_61_07	561	355	97	13
s_13_61_09	703	228	28	67
s_13_61_10	149	391	279	207

*Table 117: Item-Total Correlations and Alpha Without*

Variable	Item.Total	Alpha.Without	N
s_13_61_02	0.442	0.688	1025
s_13_61_03	0.523	0.666	1025
s_13_61_05	0.463	0.683	1025
s_13_61_06	0.367	0.710	1025
s_13_61_07	0.415	0.697	1025
s_13_61_09	0.412	0.695	1025
s_13_61_10	0.431	0.691	1025

*Table 118: Item Step Locations*

Item	nCats	Step_1	Step_2	Step_3
s_13_61_02	3.0000	-2.3966	-1.1289	1.5604
s_13_61_03	3.0000	-0.8846	0.6546	2.3494
s_13_61_05	3.0000	-0.7290	1.3250	3.2675
s_13_61_06	3.0000	-2.4407	-0.4247	1.1026
s_13_61_07	3.0000	0.1653	2.6591	5.2692
s_13_61_09	3.0000	0.6683	2.1714	3.1575
s_13_61_10	3.0000	-2.1447	0.0808	1.8102

*Table 119: IRT Item Fit (S-X2)*

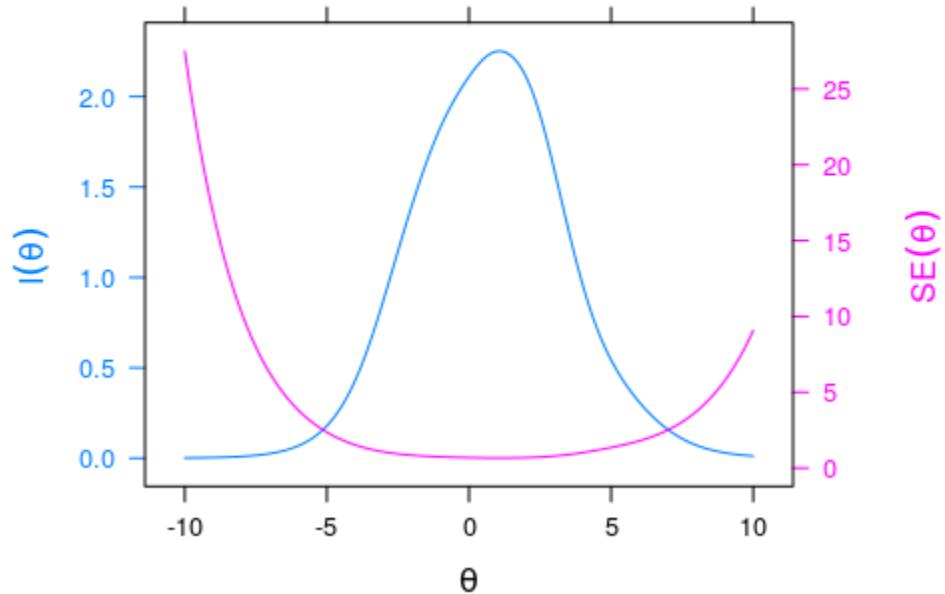
item	Fit	df	Probability
s_13_61_02	55.183	33	0.009
s_13_61_03	34.658	37	0.579
s_13_61_05	62.345	37	0.006
s_13_61_06	86.570	35	0.000

item	Fit	df	Probability
s_13_61_07	39.624	34	0.233
s_13_61_09	45.312	35	0.114
s_13_61_10	45.184	36	0.140

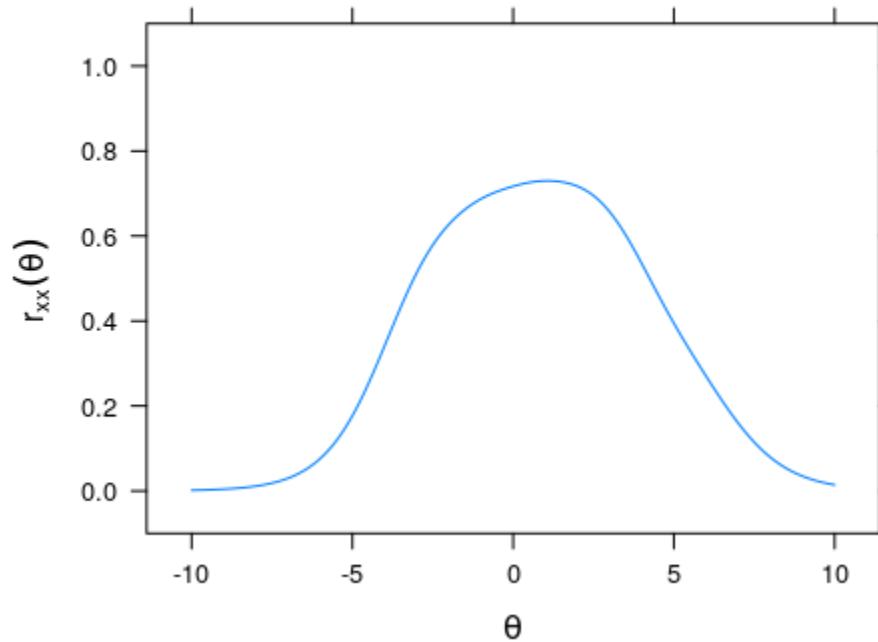
Probability < .01 indicates poor category fit

---

**Test Information (blue) and Standard Error (red)  
by Student Ability**



**Reliability by Student Ability**



### 5.30. Plate Tectonics EOU

*Table 120: Item Response Frequencies*

Item	0	1	2	3
s_13_63_02	300	297	293	123
s_13_63_03	354	93	566	
s_13_63_04	371	46	332	264
s_13_63_06	52	104	577	280
s_13_63_07	326	332	224	131
s_13_63_08	101	157	487	268
s_13_63_10	437	189	252	135
s_13_63_11	467	283	158	105
s_13_63_12	385	320	132	176

*Table 121: Item-Total Correlations and Alpha Without*

Variable	Item.Total	Alpha.Without	N
s_13_63_02	0.493	0.788	1013
s_13_63_03	0.487	0.789	1013
s_13_63_04	0.540	0.783	1013
s_13_63_06	0.421	0.797	1013
s_13_63_07	0.476	0.790	1013
s_13_63_08	0.593	0.777	1013
s_13_63_10	0.458	0.793	1013
s_13_63_11	0.543	0.782	1013
s_13_63_12	0.517	0.785	1013

*Table 122: Item Step Locations*

Item	nCats	Step_1	Step_2	Step_3
s_13_63_02	3.0000	-1.1383	0.4603	2.2972
s_13_63_03	2.0000	-0.9450	0.1942	
s_13_63_04	3.0000	-0.9672	-0.0989	1.2153
s_13_63_06	3.0000	-3.1225	-1.5811	1.0847
s_13_63_07	3.0000	-0.9787	0.7353	2.3405
s_13_63_08	3.0000	-2.1860	-0.8190	1.1030
s_13_63_10	3.0000	-0.5709	0.6223	2.0877
s_13_63_11	3.0000	-0.3671	0.9711	2.2327
s_13_63_12	3.0000	-0.6674	0.7701	1.9113

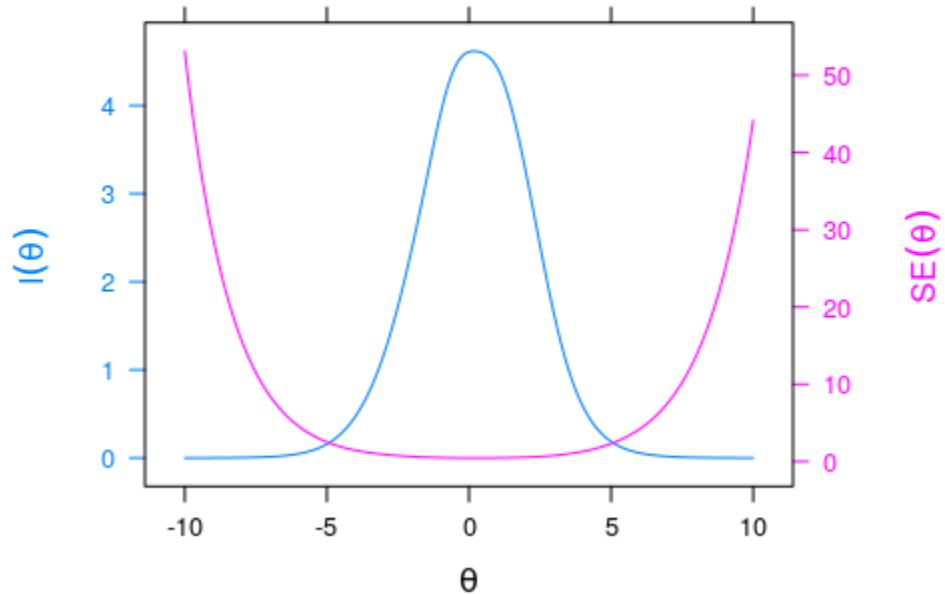
*Table 123: IRT Item Fit (S-X2)*

<b>item</b>	<b>Fit</b>	<b>df</b>	<b>Probability</b>
s_13_63_02	60.510	53	0.223
s_13_63_03	37.833	36	0.386
s_13_63_04	74.251	49	0.011
s_13_63_06	62.876	44	0.032
s_13_63_07	51.344	52	0.500
s_13_63_08	42.747	46	0.609
s_13_63_10	58.791	53	0.272
s_13_63_11	43.477	50	0.731
s_13_63_12	51.088	51	0.470

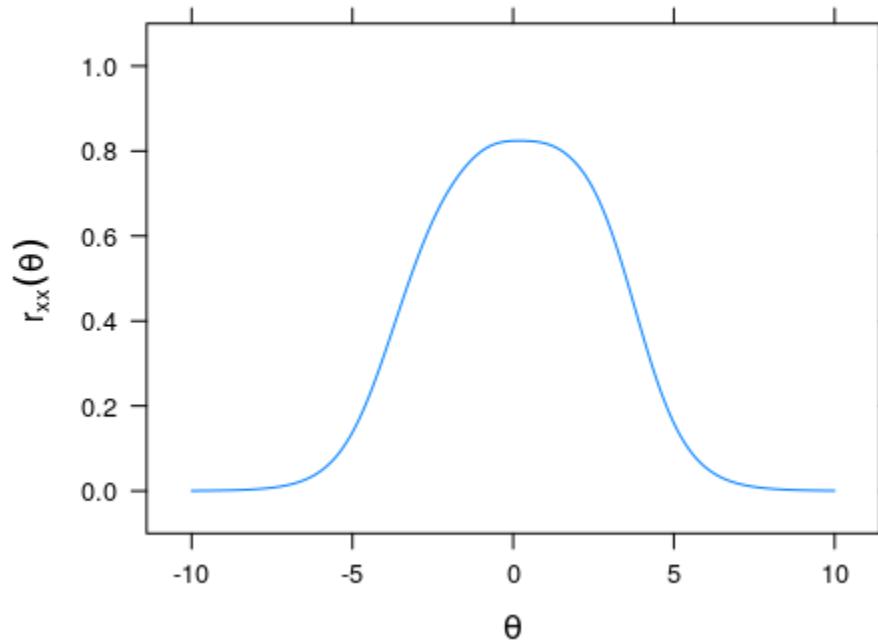
Probability < .01 indicates poor category fit

---

**Test Information (blue) and Standard Error (red)  
by Student Ability**



**Reliability by Student Ability**



### 5.31. Water on the move

*Table 124: Item Response Frequencies*

Item	0	1	2	3	4
s_14_64_02	243	287	51	76	
s_14_64_03	281	14	173	189	
s_14_64_04	117	211	215	114	
s_14_64_06	433	60	78	35	51
s_14_64_07	355	181	121		
s_14_64_09	309	153	137	58	
s_14_64_10	86	129	166	164	112

*Table 125: Item-Total Correlations and Alpha Without*

Variable	Item.Total	Alpha.Without	N
s_14_64_02	0.472	0.615	657
s_14_64_03	0.359	0.646	657
s_14_64_04	0.370	0.640	657
s_14_64_06	0.312	0.661	657
s_14_64_07	0.294	0.659	657
s_14_64_09	0.397	0.633	657
s_14_64_10	0.502	0.597	657

*Table 126: Item Step Locations*

Item	nCats	Step_1	Step_2	Step_3	Step_4
s_14_64_02	3.0000	-0.7905	1.4681	2.7369	
s_14_64_03	3.0000	-0.9125	-0.0440	1.4169	
s_14_64_04	3.0000	-2.4346	-0.0273	2.4412	
s_14_64_06	4.0000	0.3525	1.3008	2.1752	3.2034
s_14_64_07	2.0000	0.1640	2.4210		
s_14_64_09	3.0000	-0.3923	1.1270	2.9700	
s_14_64_10	4.0000	-1.9386	-0.7016	0.3309	1.6545

*Table 127: IRT Item Fit (S-X2)*

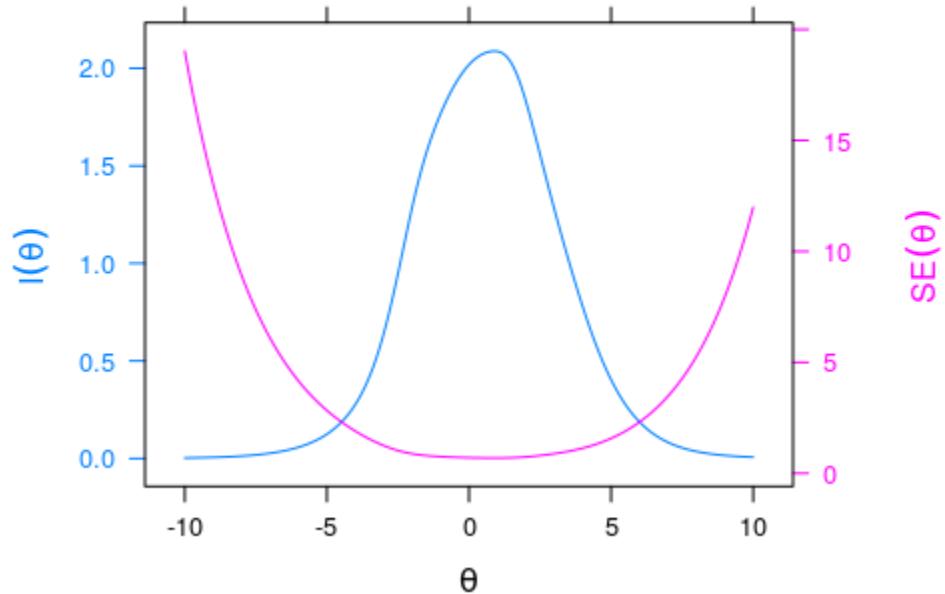
item	Fit	df	Probability
s_14_64_02	52.657	43	0.148
s_14_64_03	34.424	34	0.447
s_14_64_04	36.865	42	0.695

item	Fit	df	Probability
s_14_64_06	70.486	50	0.030
s_14_64_07	45.866	33	0.067
s_14_64_09	38.036	42	0.646
s_14_64_10	28.886	41	0.923

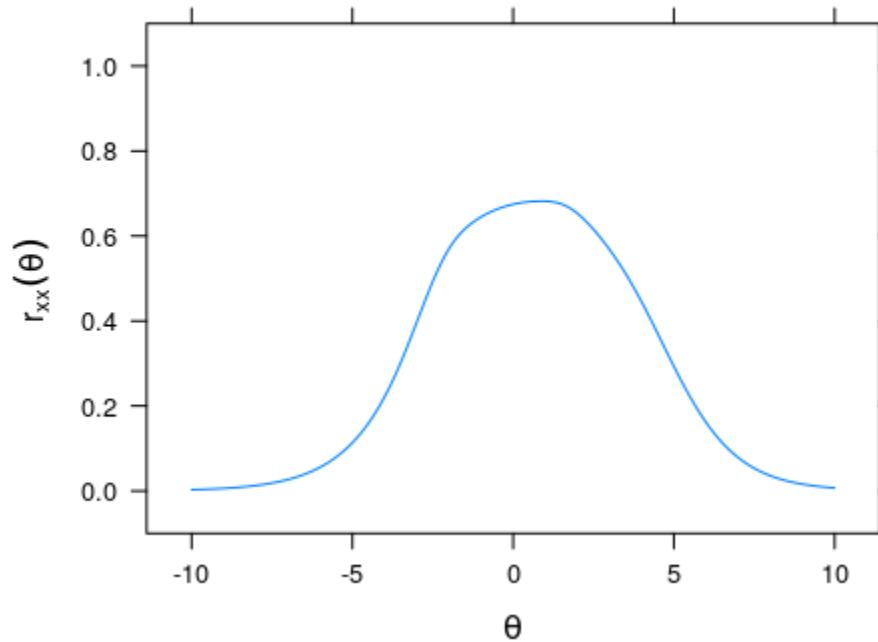
Probability < .01 indicates poor category fit

---

**Test Information (blue) and Standard Error (red)  
by Student Ability**



**Reliability by Student Ability**



## 5.32. Where's Water?

*Table 128: Item Response Frequencies*

Item	0	1	2	3
s_14_65_02	45	116	247	239
s_14_65_04	283	207	19	138
s_14_65_05	250	162	191	44
s_14_65_07	134	410	103	
s_14_65_08	305	271	63	8
s_14_65_09	260	225	97	65
s_14_65_11	84	213	334	16

*Table 129: Item-Total Correlations and Alpha Without*

Variable	Item.Total	Alpha.Without	N
s_14_65_02	0.299	0.609	647
s_14_65_04	0.423	0.567	647
s_14_65_05	0.384	0.580	647
s_14_65_07	0.246	0.621	647
s_14_65_08	0.398	0.583	647
s_14_65_09	0.347	0.593	647
s_14_65_11	0.330	0.599	647

*Table 130: Item Step Locations*

Item	nCats	Step_1	Step_2	Step_3
s_14_65_02	3.0000	-3.2914	-1.3480	0.7957
s_14_65_04	3.0000	-0.4442	0.8972	1.7613
s_14_65_05	3.0000	-0.8377	0.7624	3.1666
s_14_65_07	2.0000	-2.5896	3.1687	
s_14_65_08	3.0000	-0.2214	2.5133	5.1554
s_14_65_09	3.0000	-0.6659	1.3446	3.0165
s_14_65_11	3.0000	-2.3375	-0.0969	4.1556

*Table 131: IRT Item Fit (S-X2)*

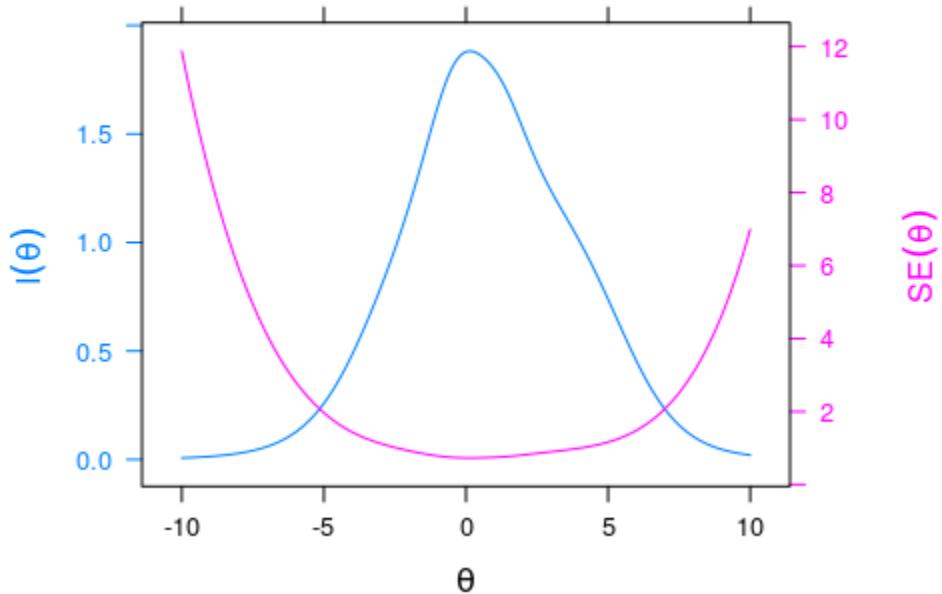
item	Fit	df	Probability
s_14_65_02	36.096	28	0.140
s_14_65_04	54.759	26	0.001
s_14_65_05	50.704	32	0.019
s_14_65_07	31.449	24	0.141

item	Fit	df	Probability
s_14_65_08	20.945	25	0.696
s_14_65_09	31.086	34	0.611
s_14_65_11	36.400	29	0.162

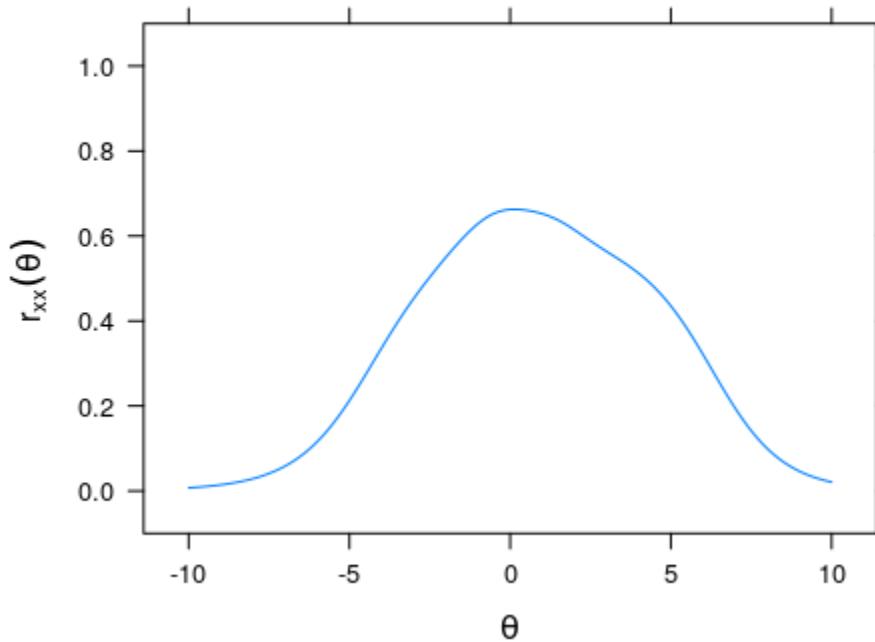
Probability < .01 indicates poor category fit

---

**Test Information (blue) and Standard Error (red)  
by Student Ability**



**Reliability by Student Ability**



### 5.33. Water And Landforms EOU

*Table 132: Item Response Frequencies*

Item	0	1	2	3	4
s_14_67_02	107	338	146	42	
s_14_67_03	247	181	61	144	
s_14_67_04	217	156	104	156	
s_14_67_06	46	158	296	133	
s_14_67_07	285	161	132	55	
s_14_67_08	60	159	180	160	74
s_14_67_09	466	130	37		

*Table 133: Item-Total Correlations and Alpha Without*

Variable	Item.Total	Alpha.Without	N
s_14_67_02	0.360	0.742	633
s_14_67_03	0.517	0.711	633
s_14_67_04	0.479	0.721	633
s_14_67_06	0.496	0.717	633
s_14_67_07	0.535	0.706	633
s_14_67_08	0.604	0.687	633
s_14_67_09	0.299	0.752	633

*Table 134: Item Step Locations*

Item	nCats	Step_1	Step_2	Step_3	Step_4
s_14_67_02	3.0000	-2.2192	1.1675	3.7095	
s_14_67_03	3.0000	-0.6761	0.5589	1.5368	
s_14_67_04	3.0000	-0.9265	0.3610	1.5250	
s_14_67_06	3.0000	-2.4348	-0.6420	1.4092	
s_14_67_07	3.0000	-0.4443	0.8697	2.3343	
s_14_67_08	4.0000	-1.9874	-0.5857	0.5232	1.9073
s_14_67_09	2.0000	1.4331	4.0351		

*Table 135: IRT Item Fit (S-X2)*

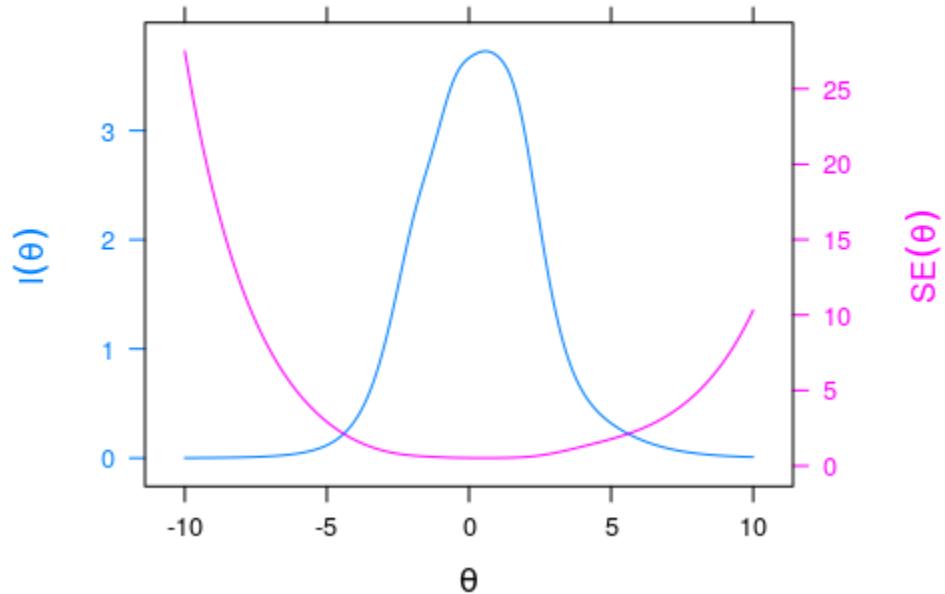
item	Fit	df	Probability
s_14_67_02	37.295	39	0.548
s_14_67_03	34.668	36	0.532
s_14_67_04	36.369	39	0.591

item	Fit	df	Probability
s_14_67_06	29.176	31	0.560
s_14_67_07	40.828	36	0.267
s_14_67_08	32.973	35	0.566
s_14_67_09	33.309	28	0.225

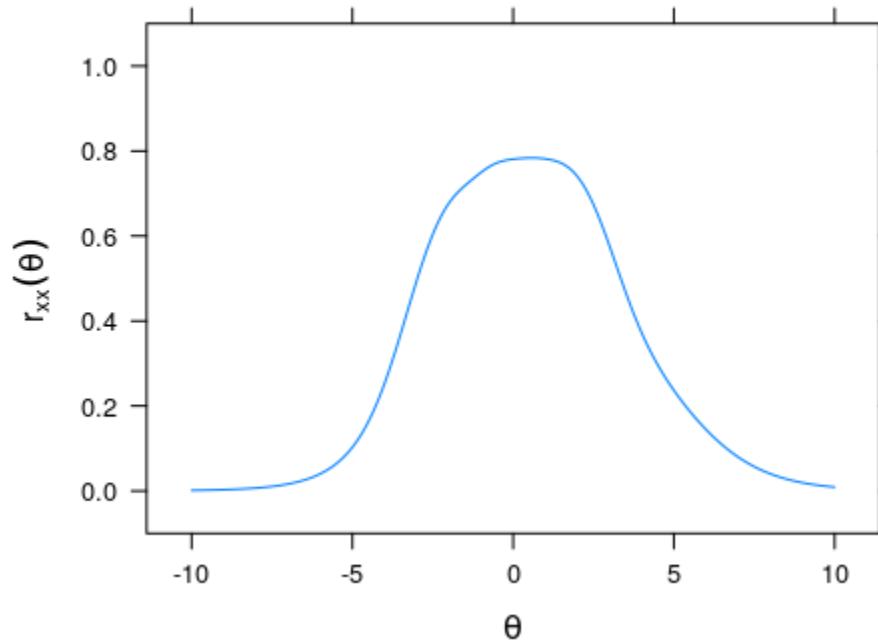
Probability < .01 indicates poor category fit

---

**Test Information (blue) and Standard Error (red)  
by Student Ability**



**Reliability by Student Ability**



### 5.34. Weather at the Beach

*Table 136: Item Response Frequencies*

Item	0	1	2	3
s_05_18_02	684	935	567	
s_05_18_03	1617	138	431	
s_05_18_05	264	1296	627	
s_05_18_06	853	594	740	
s_05_18_08	1194	491	191	311
s_05_18_11	805	1380		
s_05_18_13	1526	184	477	
s_05_18_14	1929	132	6	120

*Table 137: Item-Total Correlations and Alpha Without*

Variable	Item.Total	Alpha.Without	N
s_05_18_02	0.233	0.636	2183
s_05_18_03	0.298	0.620	2183
s_05_18_05	0.435	0.592	2183
s_05_18_06	0.302	0.620	2183
s_05_18_08	0.462	0.570	2183
s_05_18_11	0.275	0.627	2183
s_05_18_13	0.417	0.586	2183
s_05_18_14	0.302	0.618	2183

*Table 138: Item Step Locations*

Item	nCats	Step_1	Step_2	Step_3
s_05_18_02	2.0000	-1.2042	1.6512	
s_05_18_03	2.0000	1.0418	2.2372	
s_05_18_05	2.0000	-2.0485	0.9776	
s_05_18_06	2.0000	-0.7659	1.0407	
s_05_18_08	3.0000	-0.0620	1.2312	2.3046
s_05_18_11	1.0000	-0.7540		
s_05_18_13	2.0000	0.6087	1.8177	
s_05_18_14	3.0000	1.7627	2.7175	3.4567

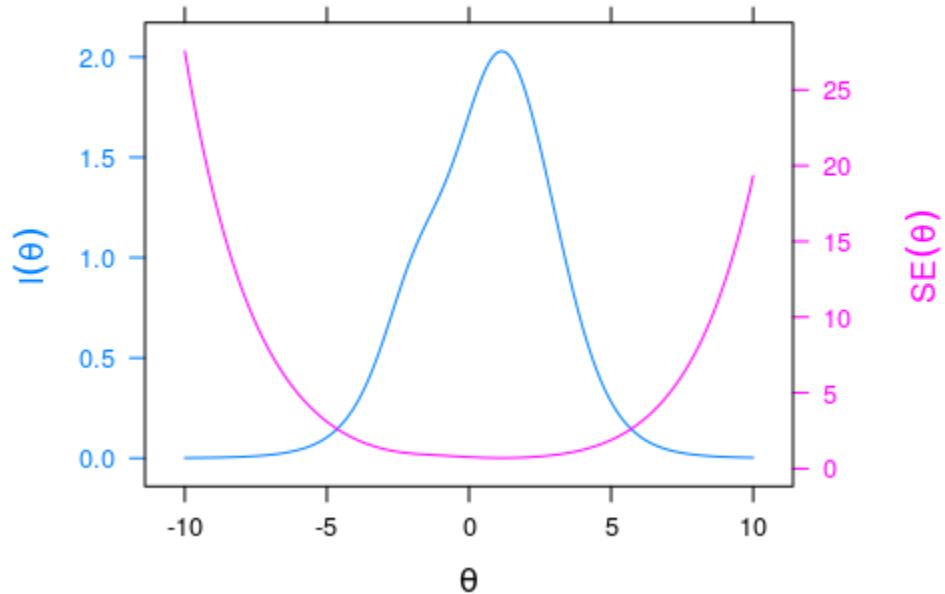
*Table 139: IRT Item Fit (S-X2)*

item	Fit	df	Probability
s_05_18_02	118.485	24	0.000
s_05_18_03	41.166	26	0.030
s_05_18_05	36.800	21	0.018
s_05_18_06	96.455	24	0.000
s_05_18_08	62.312	32	0.001
s_05_18_11	25.331	12	0.013
s_05_18_13	53.914	26	0.001
s_05_18_14	19.682	21	0.541

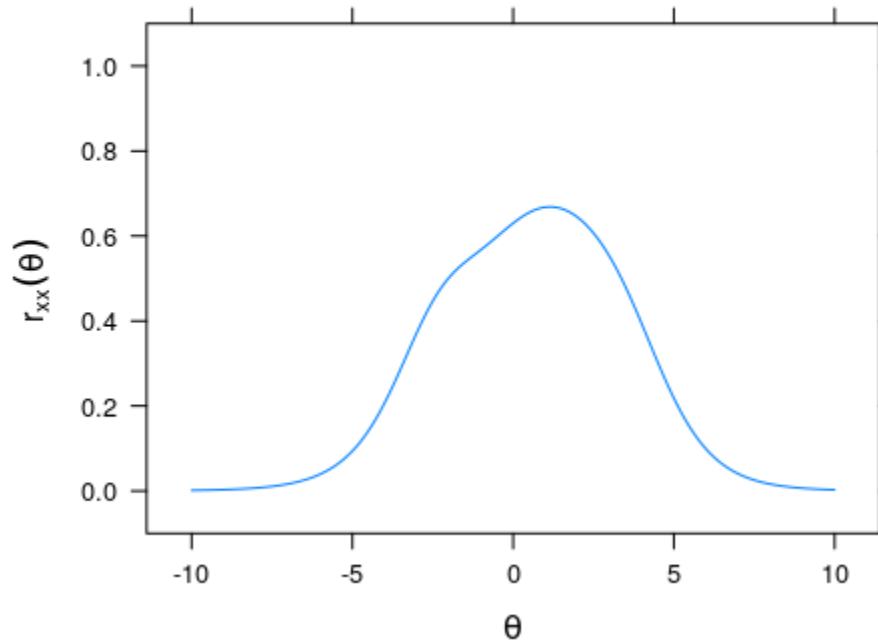
Probability < .01 indicates poor category fit

---

**Test Information (blue) and Standard Error (red)  
by Student Ability**



**Reliability by Student Ability**



## 5.35. Storms

*Table 140: Item Response Frequencies*

Item	0	1	2	3
s_05_19_02	346	889	843	
s_05_19_03	1307	572	144	54
s_05_19_05	1592	205	281	
s_05_19_06	1238	653	187	
s_05_19_08	689	1152	163	15
s_05_19_09	494	1087	497	
s_05_19_10	1285	646	146	
s_05_19_12	1257	708	108	5

*Table 141: Item-Total Correlations and Alpha Without*

Variable	Item.Total	Alpha.Without	N
s_05_19_02	0.311	0.465	2018
s_05_19_03	0.293	0.472	2018
s_05_19_05	0.167	0.521	2018
s_05_19_06	0.205	0.505	2018
s_05_19_08	0.279	0.480	2018
s_05_19_09	0.359	0.447	2018
s_05_19_10	0.262	0.486	2018
s_05_19_12	0.080	0.543	2018

*Table 142: Item Step Locations*

Item	nCats	Step_1	Step_2	Step_3
s_05_19_02	2.0000	-2.0696	0.5424	
s_05_19_03	3.0000	0.5273	2.6416	4.4050
s_05_19_05	2.0000	1.7086	3.2660	
s_05_19_06	2.0000	0.8104	4.9855	
s_05_19_08	3.0000	-0.9585	3.1501	6.4325
s_05_19_09	2.0000	-1.6598	1.5842	
s_05_19_10	2.0000	0.7459	4.1682	
s_05_19_12	3.0000	1.5784	10.0000	10.0000

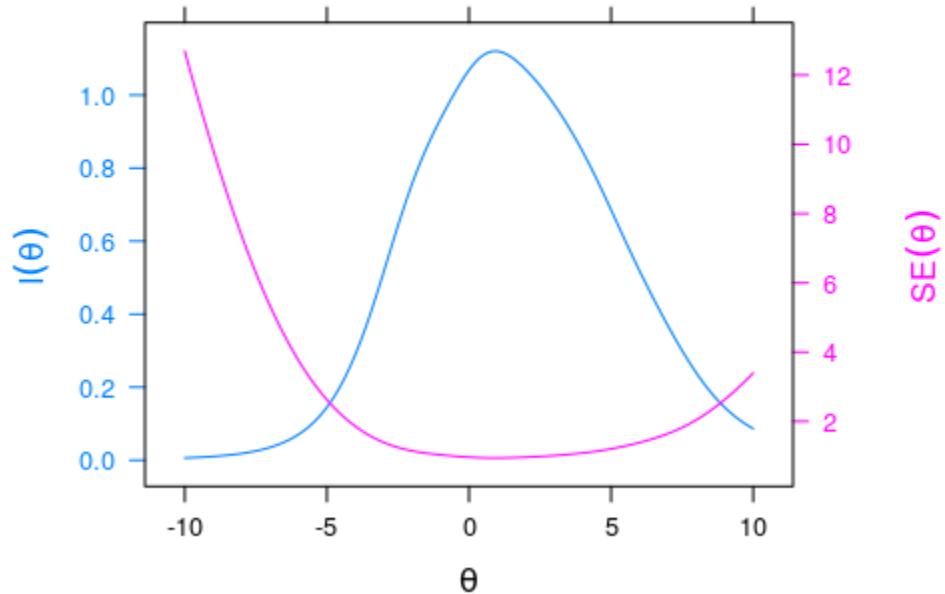
*Table 143: IRT Item Fit (S-X2)*

item	Fit	df	Probability
s_05_19_02	18.135	17	0.380
s_05_19_03	34.180	28	0.195
s_05_19_05	19.268	20	0.504
s_05_19_06	19.181	21	0.574
s_05_19_08	33.420	26	0.150
s_05_19_09	33.720	19	0.020
s_05_19_10	23.288	21	0.329
s_05_19_12	17.219	20	0.639

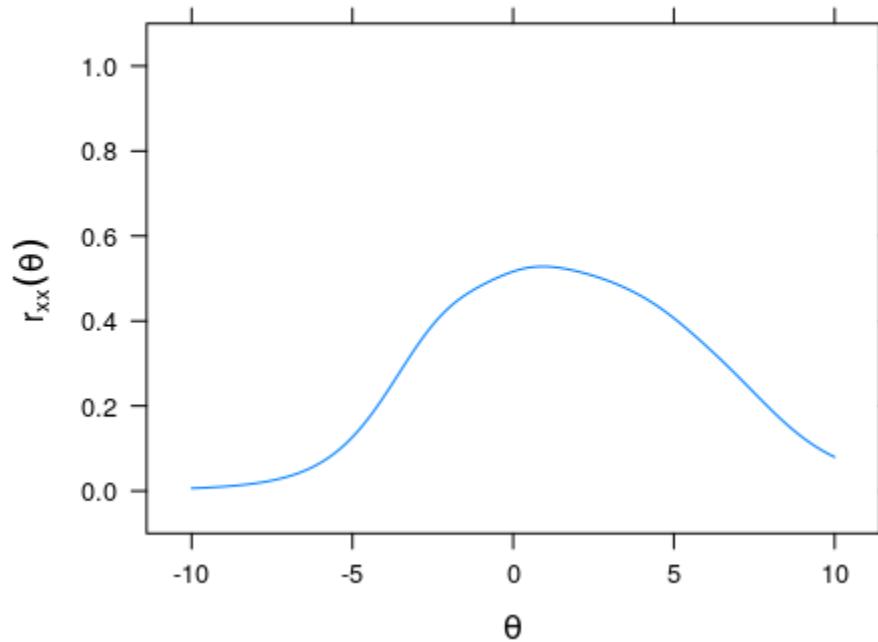
Probability < .01 indicates poor category fit

---

**Test Information (blue) and Standard Error (red)  
by Student Ability**



**Reliability by Student Ability**



### 5.36. Weather EOU

*Table 144: Item Response Frequencies*

Item	0	1	2	3
s_05_21_02	1051	212	662	
s_05_21_03	1168	756		
s_05_21_05	289	1151	346	139
s_05_21_07	605	657	614	48
s_05_21_08	1397	381	146	1
s_05_21_09	661	734	202	328
s_05_21_10	1335	554	36	

*Table 145: Item-Total Correlations and Alpha Without*

Variable	Item.Total	Alpha.Without	N
s_05_21_02	0.293	0.482	1923
s_05_21_03	0.255	0.504	1923
s_05_21_05	0.100	0.557	1923
s_05_21_07	0.308	0.473	1923
s_05_21_08	0.319	0.478	1923
s_05_21_09	0.353	0.452	1923
s_05_21_10	0.298	0.492	1923

*Table 146: Item Step Locations*

Item	nCats	Step_1	Step_2	Step_3
s_05_21_02	2.0000	-0.1342	1.1144	
s_05_21_03	1.0000	0.6332		
s_05_21_05	3.0000	-6.0000	4.5440	6.0000
s_05_21_07	3.0000	-1.1584	0.9238	4.4838
s_05_21_08	3.0000	1.3224	3.6930	6.0000
s_05_21_09	3.0000	-1.1085	1.2436	2.6269
s_05_21_10	2.0000	1.1087	5.2537	

*Table 147: IRT Item Fit (S-X2)*

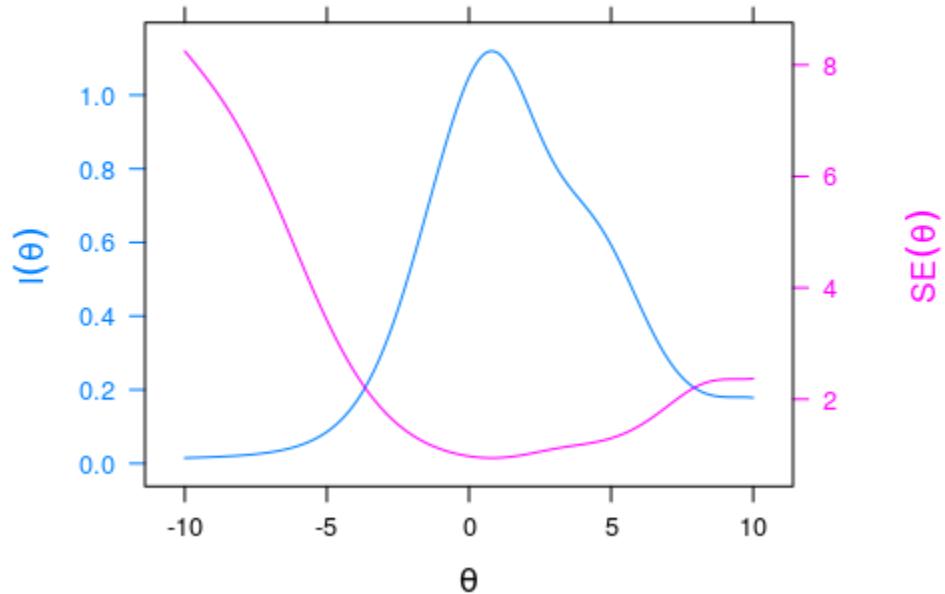
item	Fit	df	Probability
s_05_21_02	25.717	19	0.138
s_05_21_03	4.716	11	0.944
s_05_21_05	63.043	30	0.000
s_05_21_07	37.557	27	0.085

item	Fit	df	Probability
s_05_21_08	34.557	20	0.023
s_05_21_09	37.575	27	0.085
s_05_21_10	17.729	20	0.605

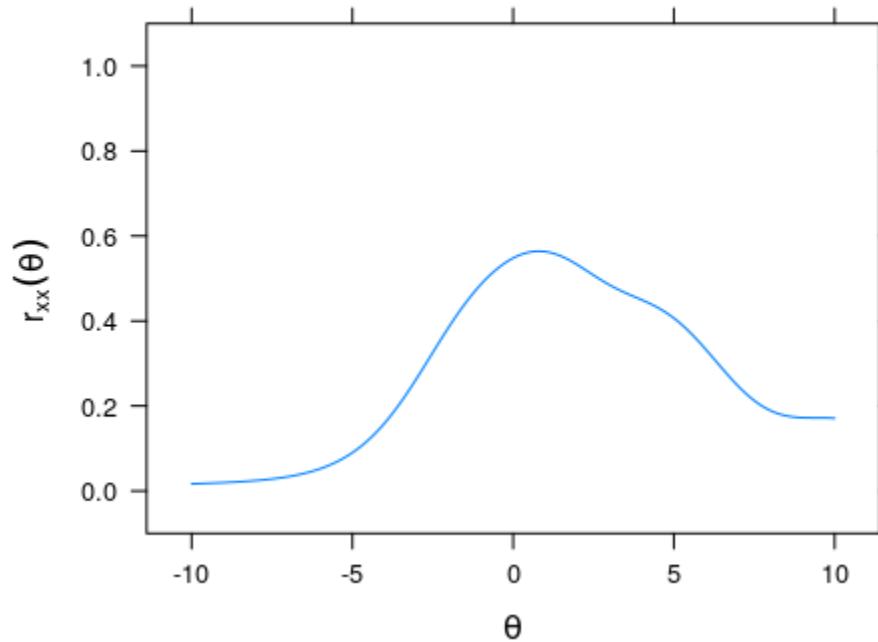
Probability < .01 indicates poor category fit

---

**Test Information (blue) and Standard Error (red)  
by Student Ability**



**Reliability by Student Ability**



## 6. Validity Studies

### 6.1. External Variables

One measure of task validity is finding that the task is related to external variables as expected. We collected information about students and teachers and compared it to student scores on each task. Table 6.1 lists the information collected including variable name, variable description, the type of variable, the level at which the variable was collected and the percent of missing values in our final dataset.

Variable	Description	Type	Source Level	pctMissing
IQWST	Teacher's IQWST Experience	ordered categorical	teacher	0
EngProf	ELL Status	ordered categorical	student	15
Teacher	Project ID for Teacher	categorical	student	0
profMath	Proficiency on State Math Assessment	logical	student	16
profELA	Proficiency on State ELA Assessment	logical	student	16
profSci	Proficiency on State Science Assessment	logical	student	82
pctNonWhite	Percent of Students with Ethnicity Other Than "White"	continuous	school	18
ses	Percent of Students categorized as Economically Disadvantaged	continuous	school	18

Note:

The variable profSci has a large percentage of missing values. Therefore, Science proficiency will be analyzed separately in a later section of this report.

### **6.1.1. Model the effects of all external factors on ONPAR task results.**

Since there are 36 separate tasks, we will analyze each one separately. To ease the interpretation of all these results, we will create an analysis of variance (ANOVA) model for one task (Plate Tectonics EOU) and see how that model applies to all tasks. We chose Plate Tectonics because it is one of the longer, more reliable tasks and as an End-of-Unit task, it covers a wider range of content than other tasks. These characteristics of the task make it representative and likely to correlate with external factors similarly to other tasks. We will start by including all external variables in our model and examine the Type II effects. This allows us to see the effect of each variable while controlling for all others. The task raw score will be the dependent variable in all our models.

TaskName	Sum Sq	Df	F value	Pr(>F)
IQWST		0		
EngProf	197.403	3	3.675	0.012
Teacher	1,037.156	3	19.307	0.000
profMath	280.899	1	15.687	0.000
profELA	350.867	1	19.595	0.000
pctNonWhite		0		
ses		0		
Teacher:profMath	717.281	6	6.676	0.000
Residuals	14,396.630	804		

We want to use only the significant effects to build an efficient model. Only the main effects of English proficiency, Teacher, Math proficiency and ELA proficiency are significant. Let's try a more efficient model including only those effects and no interactions. Note that although we did find a significant effect for the interaction between the teacher and the students' math proficiency, we couldn't come up with a reasonable meaning for that interaction in the context of this analysis. Therefore, we excluded the interaction from our efficient model.

TaskName	Sum Sq	Df	F value	Pr(>F)
EngProf	197.403	3	3.675	0.012
Teacher	7,935.100	6	73.858	0.000
profMath	280.899	1	15.687	0.000
profELA	350.867	1	19.595	0.000
Teacher:profMath	717.281	6	6.676	0.000
Residuals	14,396.630	804		

Let's see if the efficient model is as good as the original model that included all factors and their interactions. The results of an ANOVA comparison test are shown below.

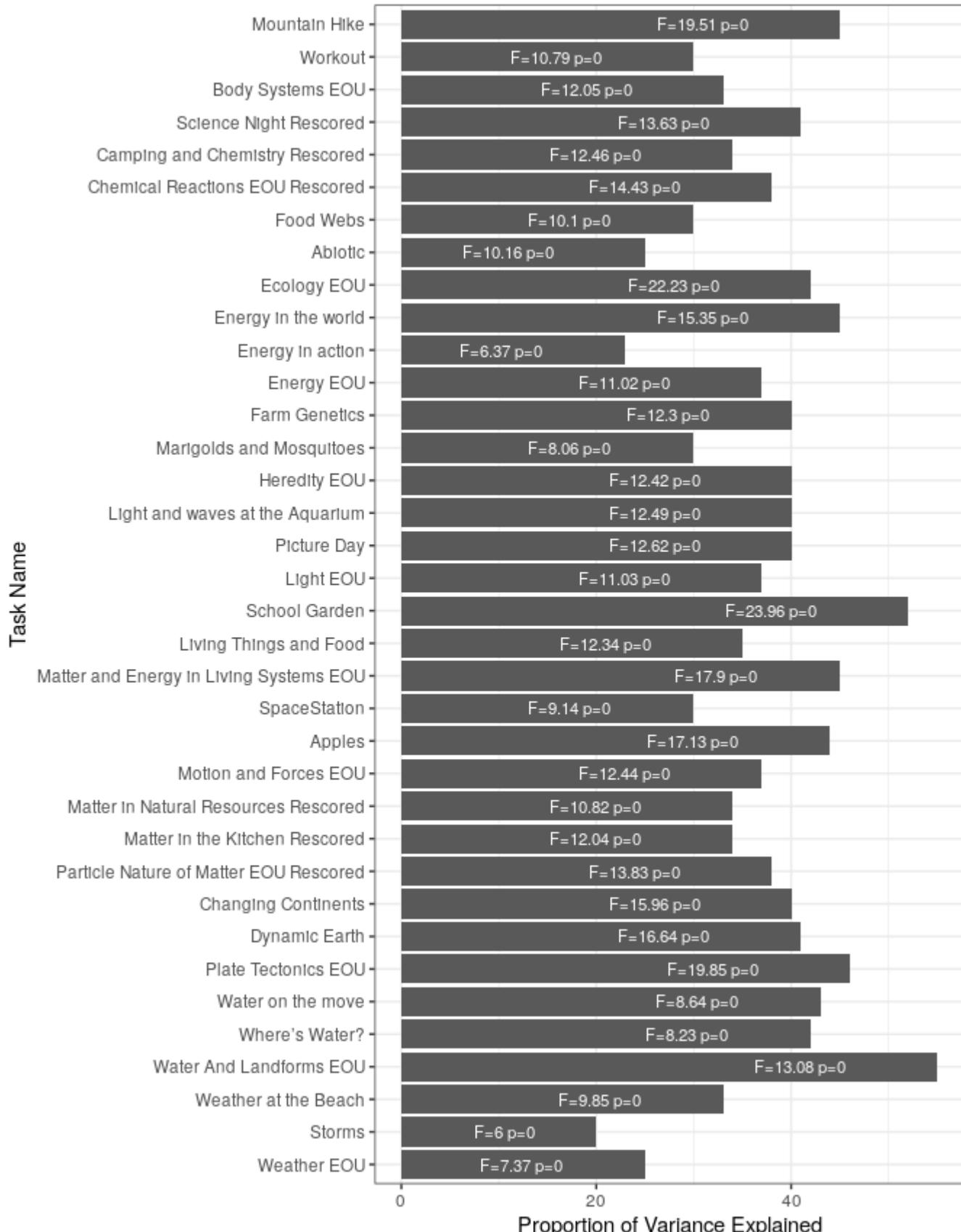
#### Analysis of Variance Table

```
Model 1: Raw ~ IQWST + EngProf + Teacher + profMath + (Teacher * profMath) +
          profELA + pctNonWhite + ses
Model 2: Raw ~ EngProf + Teacher * profMath + profELA
Res.Df   RSS Df  Sum of Sq Pr(>Chi)
1     804 14397
2     804 14397  0 -1.819e-12
```

The models are not significantly different. Therefore, it is reasonable to use the more concise model.

#### 6.1.2. Apply the Model to All Tasks.

Each row of the following figure represents the model results for each task. The F-test for the model is shown in the text in the bar. The model is significant for all tasks. The length of the bar is a display of the percent of variance in students' raw score explained by the model. We see that the model explains about 20% to 55% of the variance observed.



### 6.1.3. What is the Contribution of the Model Effects for Each Task?

Each row includes the results for a separate analysis of variance (ANOVA) for one task. The column labeled “Variance” is the adjusted R-squared \* 100 and is an estimate of the proportion of variance in the task raw score explained by all the variables in the model. Effect sizes are calculated using omega squared and interpreted using guidelines suggested by Field (2013). The right-side columns are the number of valid cases used in the analysis and the percent of valid cases compared to the total number of task scores.

TaskName	Effect Size EngProf	Effect Size ProfELA	Effect Size ProfMath	Effect Size Teacher	Variance	Cases Included	Percent Excluded
Mountain Hike	0 (Very Small)	0.09 (Medium)	0.01 (Very Small)	0.34 (Large)	45%	861	7%
Workout	0.01 (Very Small)	0.07 (Medium)	0.02 (Small)	0.19 (Large)	30%	866	7%
Body Systems EOU	0.02 (Small)	0.15 (Large)	0.02 (Small)	0.15 (Large)	33%	821	7%
Science Night Rescored	0.02 (Small)	0.11 (Medium)	0.01 (Very Small)	0.27 (Large)	41%	1073	25%
Camping and Chemistry Rescored	0.01 (Small)	0.23 (Large)	0.03 (Small)	0.07 (Medium)	34%	878	20%
Chemical Reactions EOU Rescored	0.02 (Small)	0.19 (Large)	0.04 (Small)	0.12 (Medium)	38%	878	19%
Food Webs	0.01 (Very Small)	0.15 (Large)	0 (Very Small)	0.14 (Medium)	30%	1626	64%
Abiotic	0.01 (Small)	0.16 (Large)	0 (Very Small)	0.07 (Medium)	25%	1559	67%
Ecology EOU	0.02 (Small)	0.31 (Large)	0.02 (Small)	0.08 (Medium)	42%	1517	67%
Energy in the world	0.05 (Small)	0.17 (Large)	0.02 (Small)	0.19 (Large)	45%	651	15%
Energy in action	0 (Very Small)	0.12 (Medium)	0.01 (Very Small)	0.09 (Medium)	23%	646	15%
Energy EOU	0.04 (Small)	0.22 (Large)	0.03 (Small)	0.09 (Medium)	37%	633	15%
Farm Genetics	0.03 (Small)	0.15 (Large)	0.02 (Small)	0.17 (Large)	40%	1021	5%
Marigolds and Mosquitoes	0.03 (Small)	0.15 (Large)	0.03 (Small)	0.08 (Medium)	30%	896	4%

TaskName	Effect Size EngProf	Effect Size ProfELA	Effect Size ProfMath	Effect Size Teacher	Variance	Cases Included	Percent Excluded
Heredity EOU	0.03 (Small)	0.2 (Large)	0.02 (Small)	0.15 (Large)	40%	914	4%
Light and waves at the Aquarium	0.04 (Small)	0.26 (Large)	0.01 (Small)	0.07 (Medium)	40%	914	29%
Picture Day	0.03 (Small)	0.24 (Large)	0.01 (Very Small)	0.11 (Medium)	40%	896	29%
Light EOU	0.05 (Small)	0.23 (Large)	0.01 (Very Small)	0.07 (Medium)	37%	892	30%
School Garden	0.03 (Small)	0.12 (Medium)	0.01 (Small)	0.36 (Large)	52%	943	4%
Living Things and Food	0.01 (Small)	0.1 (Medium)	0.01 (Small)	0.21 (Large)	35%	844	2%
Matter and Energy in Living Systems EOU	0.03 (Small)	0.23 (Large)	0.02 (Small)	0.17 (Large)	45%	835	2%
Space Station	0.02 (Small)	0.06 (Medium)	0.01 (Small)	0.17 (Large)	30%	1653	65%
Apples	0.05 (Small)	0.16 (Large)	0.01 (Very Small)	0.2 (Large)	44%	1918	55%
Motion and Forces EOU	0.02 (Small)	0.19 (Large)	0.02 (Small)	0.14 (Large)	37%	1649	64%
Matter in Natural Resources Rescored	0 (Very Small)	0.14 (Large)	0 (Very Small)	0.17 (Large)	34%	1032	22%
Matter in the Kitchen Rescored	0.01 (Small)	0.19 (Large)	0.01 (Very Small)	0.12 (Medium)	34%	1015	24%
Particle Nature of Matter EOU Rescored	0 (Very Small)	0.23 (Large)	0.01 (Small)	0.13 (Medium)	38%	969	22%
Changing Continents	0 (Very Small)	0.18 (Large)	0.01 (Small)	0.19 (Large)	40%	1062	16%
Dynamic Earth	0.03 (Small)	0.12 (Medium)	0.02 (Small)	0.23 (Large)	41%	1027	16%
Plate Tectonics EOU	0.01 (Small)	0.17 (Large)	0.01 (Small)	0.25 (Large)	46%	1014	16%
Water on the move	0.1 (Medium)	0.18 (Large)	0 (Very Small)	0.15 (Large)	43%	657	81%
Where's Water?	0.06 (Small)	0.18 (Large)	0 (Very Small)	0.17 (Large)	42%	647	81%
Water And Landforms EOU	0.13 (Medium)	0.21 (Large)	0.01 (Very Small)	0.18 (Large)	55%	633	81%

TaskName	Effect Size EngProf	Effect Size ProfELA	Effect Size ProfMath	Effect Size Teacher	Variance	Cases Included	Percent Excluded
Weather at the Beach	0.04 (Small)	0.13 (Medium)	0 (Very Small)	0.16 (Large)	33%	2187	34%
Storms	0.02 (Small)	0.13 (Medium)		0.05 (Small)	20%	2078	35%
Weather EOU	0.02 (Small)	0.13 (Medium)		0.1 (Medium)	25%	1925	36%

#### 6.1.3.1. Findings

- The model explains a significant proportion of the variance for all tasks.
- The total variance explained in between 20% and 55%.
- In general, the teacher and the student's proficiency on the state ELA assessments have the largest effect on task scores.
- If we consider the ELA score to be a proxy for general student academic ability, then it is unsurprising to find that teacher quality and student ability have a significant effect on student scores.
- English proficiency and proficiency on the state Math assessment have only minor effects on task scores.
- The lack of a significant Effect for English Proficiency is consistent with past research on ONPAR tasks that found our approach to assessment provided English learners a better opportunity to demonstrate their knowledge than did traditional science tests.
- Even though the ONPAR tasks measure science concepts that include math practices, math proficiency was not found to be a limiting factor in student scores.

#### 6.1.4. Science Models

As noted above, we only had enough state science assessment results for some of our tasks. After reviewing the data, we decided to include tasks with at least 175 science scores and less than 18% missing cases. This left 15 tasks to analyze. We added science proficiency to the ANOVA model and obtained the following results.

TaskName	Variance	EffSize EngProf	EffSize profELA	EffSize profMath	EffSize profSci	EffSize Teacher	Cases Included	Percent Included
Farm Genetics	47%	0 (Very Small)	0.05 (Small)	0.02 (Small)	0.1 (Medium)	0.31 (Large)	450	44%
Marigolds and Mosquitoes	33%	0 (Very Small)	0.02 (Small)	0.02 (Small)	0.11 (Medium)	0.17 (Large)	410	46%
Heredity EOU	47%	0 (Very Small)	0.06 (Small)	0.04 (Small)	0.15 (Large)	0.22 (Large)	429	47%
Light and waves at the Aquarium	43%	0.01 (Very Small)	0.08 (Medium)	0 (Very Small)	0.31 (Large)	0.03 (Small)	214	23%
Picture Day	40%	0 (Very Small)	0.05 (Small)	0 (Very Small)	0.34 (Large)	0.01 (Very Small)	209	23%
Light EOU	51%	0.02 (Small)	0.1 (Medium)	0.01 (Very Small)	0.38 (Large)	0.01 (Small)	196	22%
School Garden	53%	-0.01 (Very Small)	0.07 (Medium)	0.03 (Small)	0.15 (Large)	0.3 (Large)	230	24%
Living Things and Food	31%	-0.01 (Very Small)	0.02 (Small)	0.03 (Small)	0.15 (Large)	0.12 (Medium)	190	23%
Matter and Energy in Living Systems EOU	55%	-0.01 (Very Small)	0.06 (Medium)	0.02 (Small)	0.22 (Large)	0.24 (Large)	191	23%
Space Station	29%	0 (Very Small)	0 (Very Small)	0.01 (Very Small)	0.09 (Medium)	0.19 (Large)	284	17%
Apples	42%	0.01 (Very Small)	0.09 (Medium)	0.02 (Small)	0.16 (Large)	0.14 (Large)	277	14%
Motion and Forces EOU	51%	0 (Very Small)	0.1 (Medium)	0.03 (Small)	0.14 (Medium)	0.24 (Large)	292	18%
Changing Continents	42%	0 (Very Small)	0.14 (Large)	0.02 (Small)	0.25 (Large)	0.02 (Small)	195	18%
Dynamic Earth	27%	0 (Very Small)	0.04 (Small)	0.06 (Small)	0.17 (Large)	0 (Very Small)	190	19%

TaskName	Variance	EffSize EngProf	EffSize profELA	EffSize profMath	EffSize profSci	EffSize Teacher	Cases Included	Percent Included
Plate Tectonics EOU	34%	0.02 (Small)	0.06 (Medium)	0.02 (Small)	0.24 (Large)	0 (Very Small)	180	18%

#### 6.1.4.1. Findings

- The model explains a significant proportion of the variance for all tasks. Across the 15 tasks, the model explained 27% - 55% of the variance in raw scores.
- Science proficiency had a consistently large effect across tasks indicating that ONPAR is measuring the science standards.
- Since state proficiency tests are moderately correlated, the science proficiency effect may include general student academic ability.
- The teacher effect remained large for most tasks.
- The effects of English proficiency and proficiency on the state Math assessment continued to be negligible.
- ELA proficiency is no longer a large effect after science proficiency added.

Field, A. (2013) *Discovering statistics using IBM SPSS Statistics*. Fourth Edition. Sage: London.