

---

# Automated Test Assembly (ATA)

**Case Studies in  
Classical Test Theory (CTT) and  
Item Response Theory (IRT)**

**November 3, 2017**

*David Chuah*

*Luz Bay*

*Thomas Proctor*

The College Board

# How to assemble a test form

Imagine if you have to pick 20 items for a test form

- Examples of Hard Constraints
  - Math
  - Reading
  - Multiple Choice/ Free Response
  - Difficulty
  - Multiple Forms
- Layers
- Soft Constraints

Constraint	Hard Low	Soft Low	Soft High	Hard High
Math	10			10
MCQ	3	4	5	6
Free Response		5	6	

# Automated Test Assembly (ATA)

- Computer program to select items for a test
  - Linear test form
    - Linear-on-the-fly (LOFT)
    - Fixed length form
  - Multi-stage test
  - Computer Adaptive Test (CAT) ~ Shadow Test
- Linear Programming/Optimization  
Mixed Integer Programming is LP restricted to integers
- College Board currently uses ATA for
  - SAT suite of assessments,
  - ACCUPLACER®
    - COMPANION Form
    - Standard Setting Ordered Item Booklet (OIB)

---

For ATA to work you need

- 
- Item Pool with item metadata
    - Content
    - Item Statistics
  - Test specifications/Constraints
    - Number of items on a test
    - Number of items by Content
    - Difficulty profile of the test items
  - ATA Software

---

# ATA Types and Strategies

---

## Types

- Heuristic
- Linear Optimization

## Build Strategy

- Sequential Building
- Simultaneous Building

---

# Linear Programming Solvers

- 
- LP Solve – free Mixed Integer Linear Programming (MILP)
  - GLPK – free GNU Linear Programming package
  - Symphony – Open-source MILP
  - Gurobi – commercial : linear/quadratic/mixed integer
  - FICO Xpress – commercial
  - IBM ILOG CPLEX - commercial

# ATA Example 1

Example provided by Donovan Hare

Hibachi Model	X	Y	Maximum per day
Profit	\$2.00	\$1.50	
Cast Iron Ingots	3	4	1000
Labor Units	6	3	1200

$$\text{Profit} = 2.00x + 1.50y$$

$$3x + 4y \leq 1000$$

$$6x + 3y \leq 1200$$

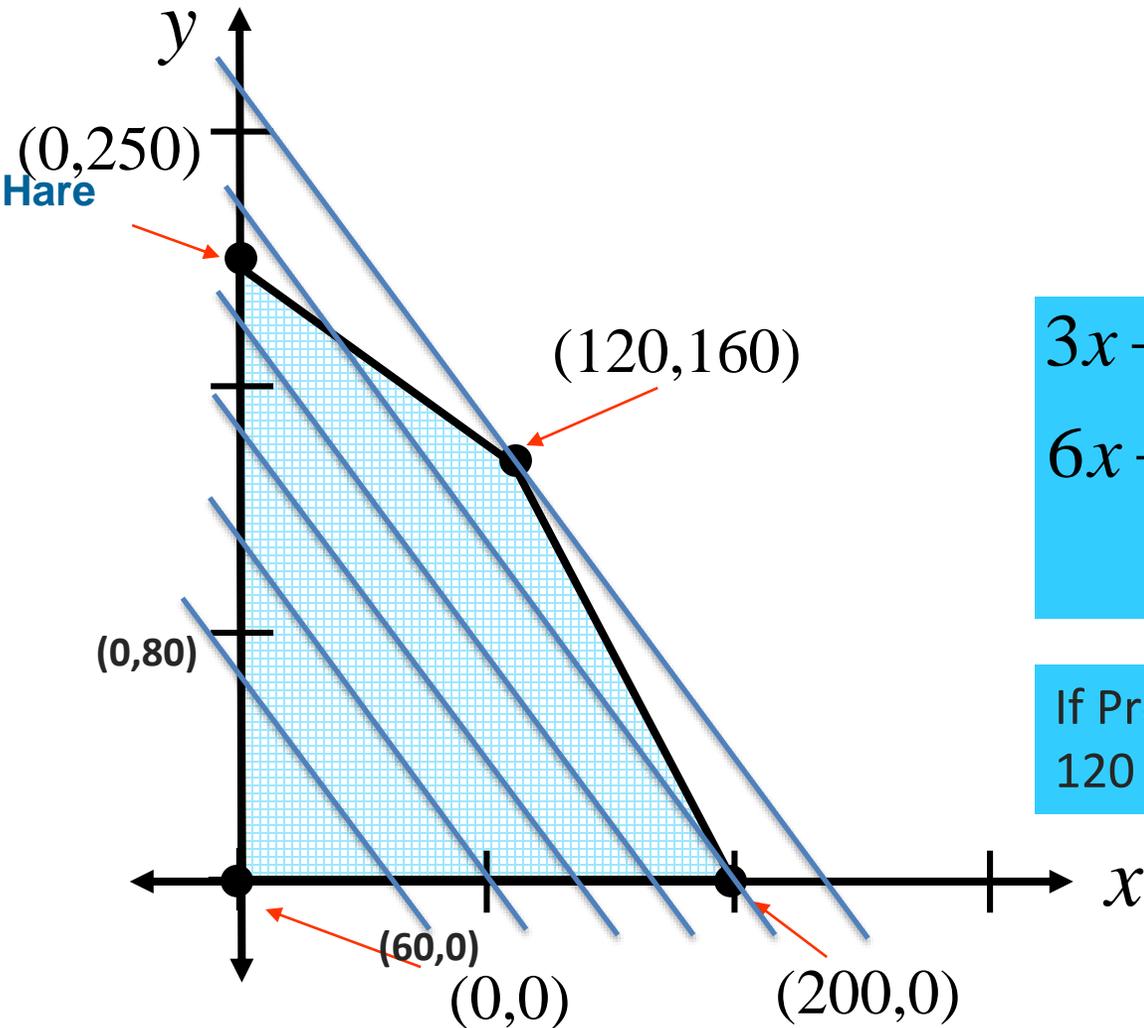
$$x \geq 0, y \geq 0$$

## Limitations

- Linear objective  
(No division or multiplication)
- Cannot be strict inequality  
(No '<' or '>')

# ATA Example 1 (continued)

Example provided by Donovan Hare



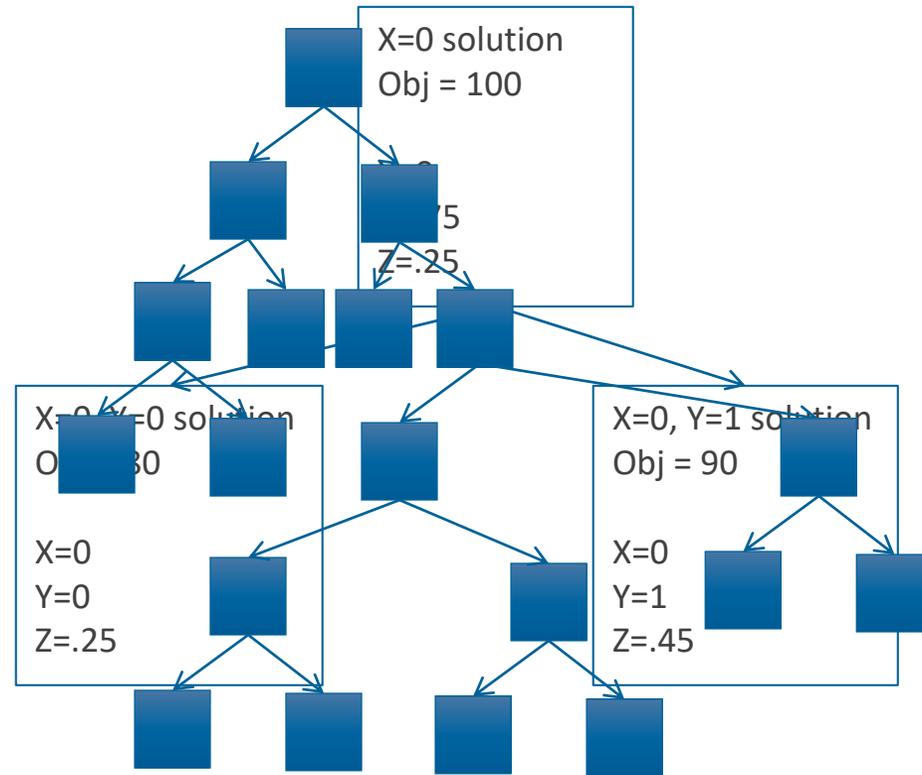
$$\begin{aligned} 3x + 4y &\leq 1000 \\ 6x + 3y &\leq 1200 \\ x, y &\geq 0 \text{ integer} \end{aligned}$$

$$\begin{aligned} \text{If Profit} &= \$120 \\ 120 &= 2x + 1.5y \end{aligned}$$

# ATA Example 2

## Search

### Branch and Bound



---

# Example of ATA Software

IBM ILOG CPLEX

- 
- Reading in Data
  - Test constraint parameters & definitions
  - Optional: Time or Gap limit
  - Linear Optimization Function  
(maximize or minimize)
  - Express constraints
  - Output results

---

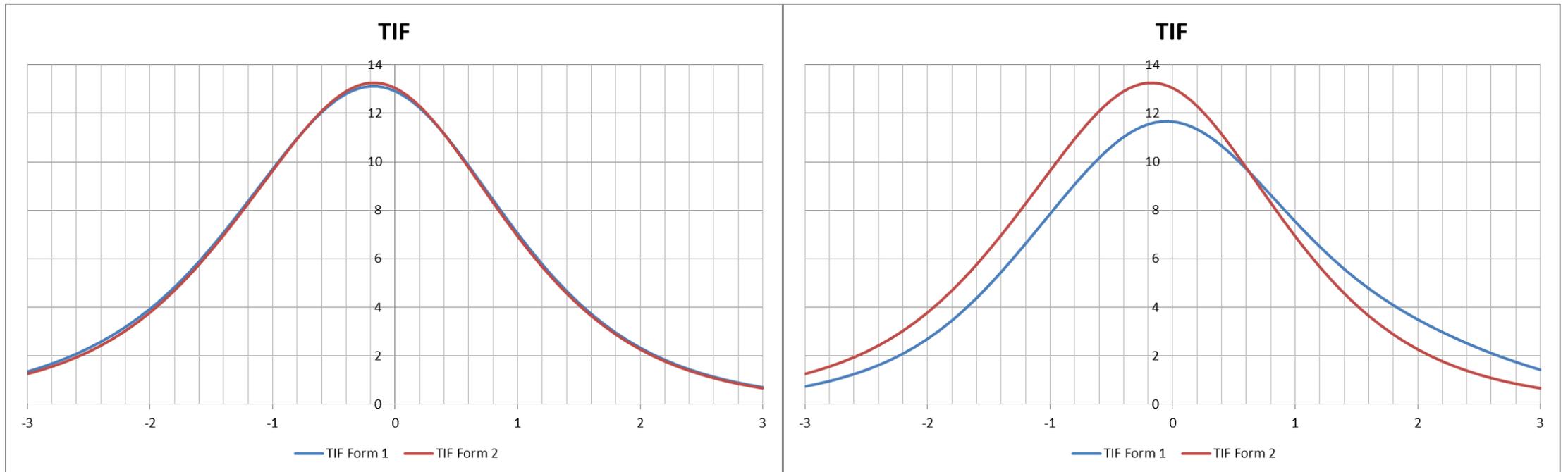
# ACCUPLACER

## Case Study in IRT

- 
- Fully adaptive CAT
  - Three-parameter logistic (3PL) model
  - Application 1: COMPANION Forms  
Paper-based linear version (i.e., not adaptive)
  - Application 2: OIB for standard setting  
Subset of item pool that matches test constraints

## Application 1: COMPANION Forms

- Simple
  - Optimization Programming Language – OPL Code
  - Excel File – Item Metadata



---

## Application 1: COMPANION Forms (continued)

---

### Possible test constraints

- # of items
- # overlap items
- TIF higher than target
- Parallel TIF
- Maximize TIF overall (and/or at cut)
- Item Sets
- Content

---

## Application 2: Ordered Item Booklet

- 
- Simple
    - Optimization Programming Language – OPL Code
    - Excel File – Item Metadata
  - RP67

---

## Application 2: Ordered Item Booklet (continued)

---

### Constraints

- # of items
- Item can only appear once
- Alternate items by content type
- Space out items evenly across difficulty
- Content

(Manual adjustment)

---

# SAT Suite of Assessments

## Case Study in the CTT Context

- 
- Paper-based linear tests (i.e., not adaptive)
  - Classical Test Theory
    - Equated p-values
    - Reliability
  - Tests
    - Reading
    - Writing
    - Math
    - Cross-Scores
    - Sub-Scores

---

## Application 3: SAT Suite of Assessments Forms

- 
- Differences between ATA implementations in the CTT and IRT contexts
  - College Board Proprietary ATA Engine
  - Data Management
  - Quality Control
  - Archival/Documentation
  - Troubleshooting
  - Multiple parallel test forms
  - Issues with assembly efficiency/speed

# Differences between CTT and IRT

## Test Reliability

### CTT

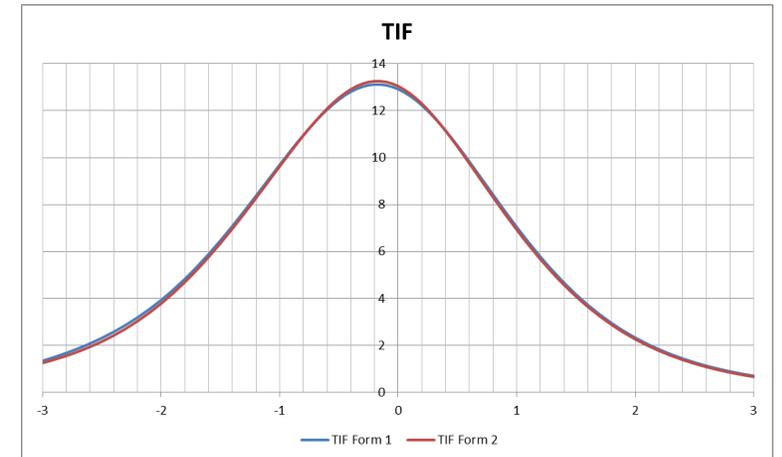
- Point biserial correlation
- Constraint numerator and denominator for reliability coefficient (van der Linden 2005)

$$\alpha = \frac{n}{n-1} \left[ 1 - \frac{\sum_{i=1}^n \sigma^2}{\left( \sum_{i=1}^n \sigma_i \rho_i x \right)} \right]$$

- Quadratic optimization
- Hybrid of quadratic and linear optimization

### IRT

- Test Characteristic Curve
- Test Information Function



# Differences between CTT and IRT

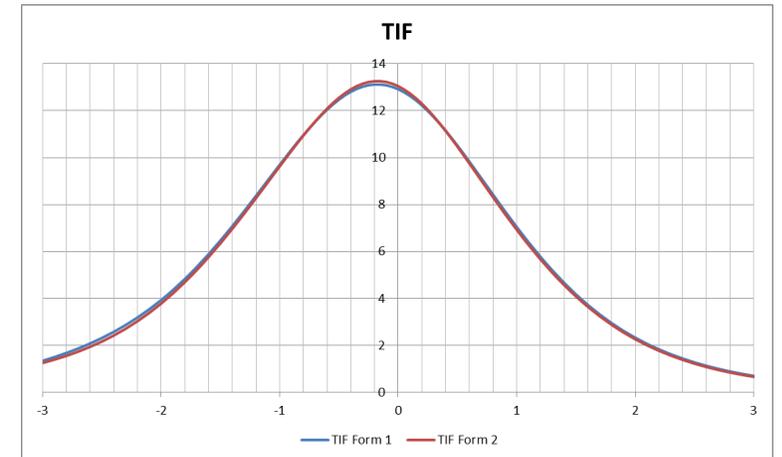
## Test Difficulty

### CTT

- Average Equated P-value/Equated Delta

### IRT

- Test Characteristic Curve
- Test Information Function

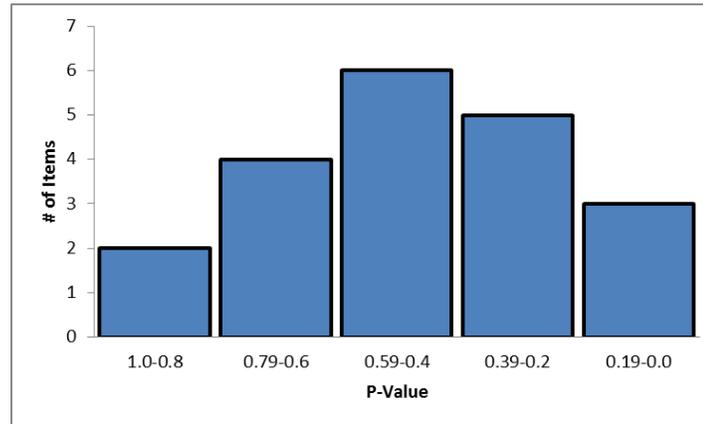


# Differences between CTT and IRT

## Distributing Test Items by Difficulty

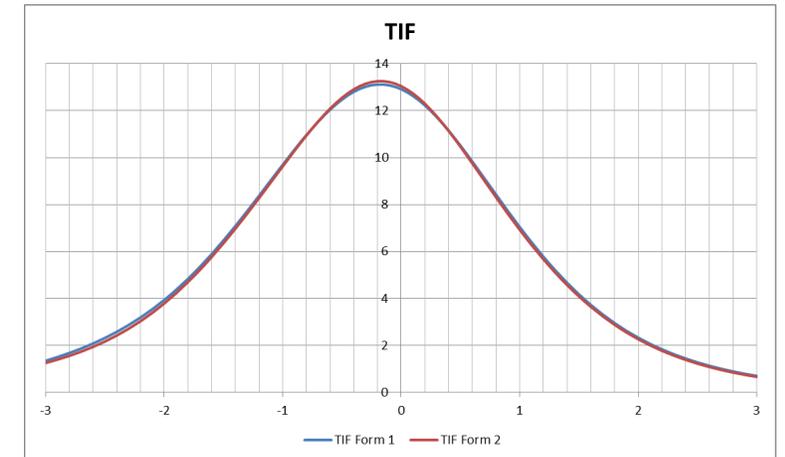
### CTT

- Bin method for Equated P-value/Equated Delta
- Distance between P-value/Delta values
- Standard deviation of P-value/Delta using quadratic optimization



### IRT

- Test Characteristic Curve
- Test Information Function



---

# College Board Proprietary ATA Engine

- 
- General Purpose ATA Program
    - Apply to new tests by changing specifications
    - SAS produces standardized files that are processed by the engine
  - Discrete & Item Sets
  - Complex Dependencies
  - CTT & IRT

---

# Aspects of Implementing ATA

- 
- Data Management
    - Excel file with test constraints
    - Constraints was designed to look similar to the files provided to us by test development
  - Quality Control
    - Blueprint highlights
    - Independent files
  - Archival/Documentation
    - Record of each build step
    - Memorandum of build

---

# Aspects of Implementing ATA (continued)

- 
- Troubleshooting
    - Debug
    - Turn-to-Soft
  - Multiple parallel test forms
    - Simultaneous solutions
    - Sequential mixed builds
  - Assembly Efficiency
    - Warm starts
    - Improved algorithms
    - Hybrid algorithm for reliability: quadratic and linear

---

## Other ATA Issues

- 
- Exposure control
  - Item Inventory Prediction

---

# Thank You.

[dchuah@collegeboard.org](mailto:dchuah@collegeboard.org)