Small-Sample Shadow Testing

Wallace Judd, Ph.D.
President,
Authentic Testing Corporation
WJudd@AuthenticTesting.com
Shadow Testing

... is a generalized constraint-satisfaction algorithm.
Shadow Testing

... is computerized adaptive test in that it adapts to the constraints being resolved...

Not necessarily to examinee ability.
Shadow Testing

Technical Explication

Qi Diao
Hao Ren

Optimal Solution to Constraints

vs

Sufficing Solution to Constraints
Shadow Testing Takeaways ...  

You can do shadow testing, successfully,  

• Without mathematical formalisms.
Shadow Testing Takeaways ...

You can do shadow testing, successfully,

• Without mathematical formalisms
• With relatively small calibration samples
Shadow Testing Takeaways ...

You can do shadow testing, successfully,

• Without mathematical formalisms
• With relatively small calibration samples
• With small item pools
Shadow Testing Takeaways …

You can do shadow testing, successfully,

• Without mathematical formalisms
• With relatively small calibration samples
• With small item pools
• Using a pseudo-information function
Shadow Testing Takeaways ...

You can do shadow testing, successfully,

- Without mathematical formalisms
- With relatively small calibration samples
- With small item pools
- Using a pseudo-information function
- To create unique equivalent test forms
  ... for each examinee
This Case:

... involves a client with a very common set of constraints

Frequently encountered...

That Shadow-Testing resolved.
# Constraints: CAT vs. Shadow

<table>
<thead>
<tr>
<th>Constraint</th>
<th>CAT</th>
<th>Shadow-Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goal</td>
<td>Measure Ability</td>
<td>Pass/Fail</td>
</tr>
<tr>
<td>Maximize</td>
<td>Precision</td>
<td>Equivalence</td>
</tr>
<tr>
<td>Length</td>
<td>Variable</td>
<td>Fixed</td>
</tr>
<tr>
<td>Stopping Rule</td>
<td>Posterior Est.</td>
<td>N of Items</td>
</tr>
<tr>
<td>Domains</td>
<td>No</td>
<td>Multiple</td>
</tr>
<tr>
<td>Item Exposure</td>
<td>Insignificant</td>
<td>Critical</td>
</tr>
</tbody>
</table>
## Constraints: CAT vs. Shadow

<table>
<thead>
<tr>
<th></th>
<th>Large</th>
<th>Small: 3:1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item Pool</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calibration N</td>
<td>500+</td>
<td>&lt;50</td>
</tr>
<tr>
<td>Constraints</td>
<td>One</td>
<td>5:1</td>
</tr>
</tbody>
</table>

- Domain Count = Difficulty
- Exposure
- Cognitive Level ≈ Time
Test Construction

Equivalent Difficulty
Multiple (7) Domains
Fixed Length (41 items)
Pass / Fail Result
Test Construction

Conditions:

Calibration Sample: . . 30 !!!
Annual Tests: . . . . . . 200-300
Item Pool: . . . . . . . . 120 Items
Domains: . . . . . . . . . 7
Items Administered. . 41
Test Construction

Constraint #1:

Draw Items from Domains
as specified in Test Blueprint
Test Construction

Classical Test Theory:

\[ P\text{-Val}_i = \text{Probability Correct Response, for Item } i \]

Constraint #2:

Minimize: \( \mu \text{ P-Val} - \text{Target P-Val} \)

Acceptable: \( \mu \text{ P-Val} - \text{Target P-Val} \leq 0.04 \)
Test Construction

Constraint #3:

Minimize item exposure
Test Construction

Constraint #4:

Match Blueprint for Item Cognitive Level
Test Construction

Constraint #5:

Create forms of equivalent expected Item Latency
Test Construction

Constraint #5:
Create forms of equivalent expected Item Latency
Test Construction

First Attempt:

Program it all as a set of conditions solved in multiple passes.
Test Construction

First Attempt:
Program it all as a set of conditions solved in multiple passes.

What A MESS!
Test Construction Problems:

Constraint #1:  
Draw Items from Domains  
as specified in Test Blueprint

Some domains had few items  
over and above the minimum.
Test Construction Problems

Constraint #2:

Minimize: $\mu \text{ P-Val} - \text{Target P-Val}$

Acceptable: $\mu \text{ P-Val} - \text{Target P-Val} \leq 0.04$

What about discriminating power?
Some items were always the best.
Test Construction Problems

Constraint #3:
Minimize item exposure

Way too exacting.
Test Construction Problems

Constraint #4:
Match Blueprint for Item Cognitive Level

Easily satisfied (except in small domains) since there are only 2 Levels
Test Construction Problems

Constraint #5:

Create forms of equivalent expected Item Latency

The tail wagging the dog...

Often unsatisfied.
Test Construction Answers:

Constraint #1:  
Draw Items from Domains  
as specified in Test Blueprint

Start with small domains, or  
one with a small Item: Target N ratio.
Test Construction Answers:

Constraint #1:

Draw Items from Domains
as specified in Test Blueprint

Randomize item seeding for
initial 10 items... from small domains.
Test Construction Answers:

Item Domain Spec in Blueprint

#1
#2
#3
#4
#5
#6
#7
Test Construction Answers:

Domain Spec & Available Items

#1  #2  #3  #4  #5  #6  #7
Test Construction Answers:

Domain Spec & Available Items

#5 #7 #4 #3 #1 #6 #2
Test Construction

Constraint #2:

Minimize: $\mu \text{ P-Val} - \text{Target P-Val}$

Acceptable: $\mu \text{ P-Val} - \text{Target P-Val} \leq 0.04$

Pseudo-Information function drawn from Classical Test Theory statistics
pInfo (pseudo-Information)

Classical Test Theory statistics:

\[ P-\text{Val}_i = \text{Probability Correct Response, for Item } i \]

\[ \text{PBis}_i = \text{Point-Biserial, Item } i \]

\[ \text{pInfo}_i = \text{PBis}_i + 1 - (\text{ABS} [ \text{Cutpoint} - P-\text{Val}_i ]) \]
Test Construction

Constraint #3:
Minimize item exposure

Relax constraint. Only evaluate when item exposure > 5 exposures out of line.
Then take out of pool.
Test Construction

Constraint #4:

Match Blueprint for Item Cognitive Level

Easily satisfied (except in small domains).

Set target as ratio of 2:1 Tasks : Knowledge, with +/- 15% sufficient.
Test Construction

Constraint #5:
Create forms of equivalent expected Item Latency

Evaluate $\Sigma$ Latency as Constraint #3.
Test Construction

Construct test form prior to administration.

If form doesn’t resolve, try again.

Yield success: attempts $\approx 1 : <3$

Form equivalence
Test Construction Results

• Domain count consistent with Blueprint
Test Construction Results

- Domain count consistent with Blueprint
- \( \mu \text{ P-Val} – \text{Target P-Val} < 0.04 \)
Test Construction Results

• Domain count consistent with Blueprint
• $\mu$ P-Val – Target P-Val < 0.04
• S Latency – Target Response Time < 5.0 min.
Test Construction Results

- Domain count consistent with Blueprint
- $\mu$ P-Val – Target P-Val < 0.04
- S Latency – Target Response Time < 5.0 min.
- Cognitive Level – Target Level = +/- 0.15
Test Construction Results

- Domain count consistent with Blueprint
- $\mu$ P-Val – Target P-Val < 0.04
- S Latency – Target Response Time < 5.0 min.
- Cognitive Level – Target Level = +/- 0.15

Q. E. D.
Shadow Testing Takeaways ...

In doing Shadow-Testing with Small N Samples

- Seed item selection with randomization
- Seed small domains first
- Use a pseudo-information function to integrate difficulty and discrimination
- Incorporate $\Sigma$ item time in targets
Test Construction

A Sufficing Solution

... Inspired by Shadow-Testing

... with apologies to Wim van der Linden
Small-Sample Shadow Testing

Wallace Judd, Ph.D.
President,
Authentic Testing Corporation
WJudd@AuthenticTesting.com