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

Constraint	CAT	Shadow-Test
Goal	Measure Ability	Pass/Fail
Maximize	Precision	Equivalence
Length	Variable	Fixed
Stopping Rule	Posterior Est.	N of Items
Domains	No	Multiple
Item Exposure	Insignificant	Critical

Constraints: CAT vs. Shadow

Item Pool	Large	Small: 3:1
Calibration N	500+	<50
Constraints	One	5: 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-Val$_i$ = Probability Correct Response, for Item i

Constraint #2:

Minimize: μ P-Val – Target P-Val

Acceptable: μ P-Val – Target P-Val <= 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 tests 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: μ P-Val – Target P-Val

Acceptable: μ P-Val – Target P-Val <= 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

ones with a small Item: Target N ratio.

Test Construction Answers:

Constraint #1:

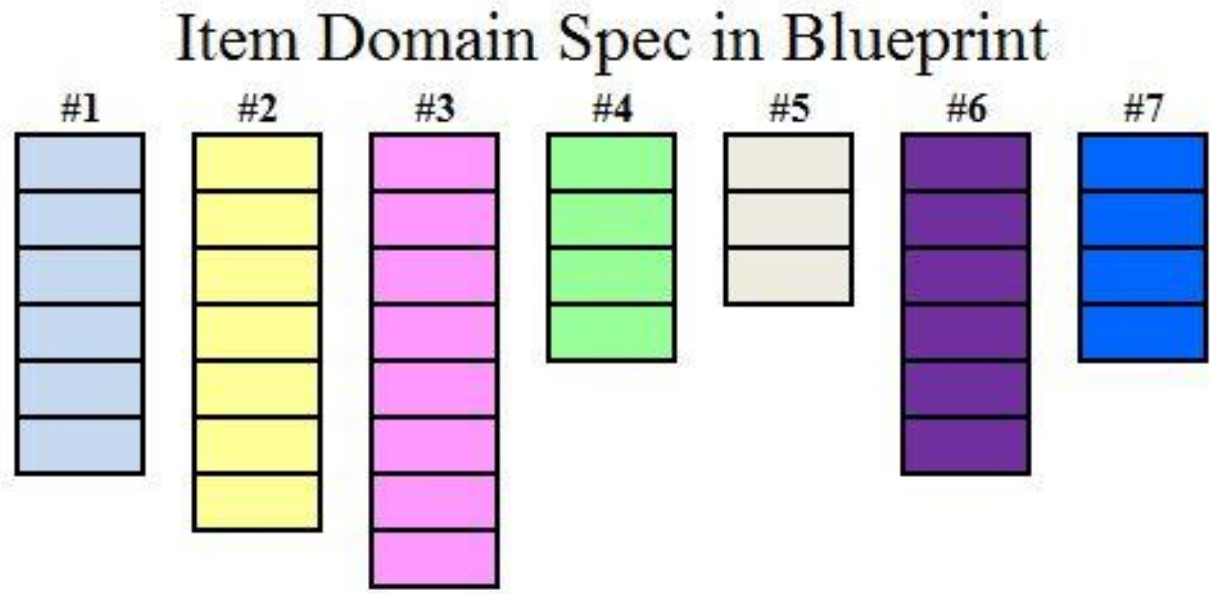
Draw Items from Domains

as specified in Test Blueprint

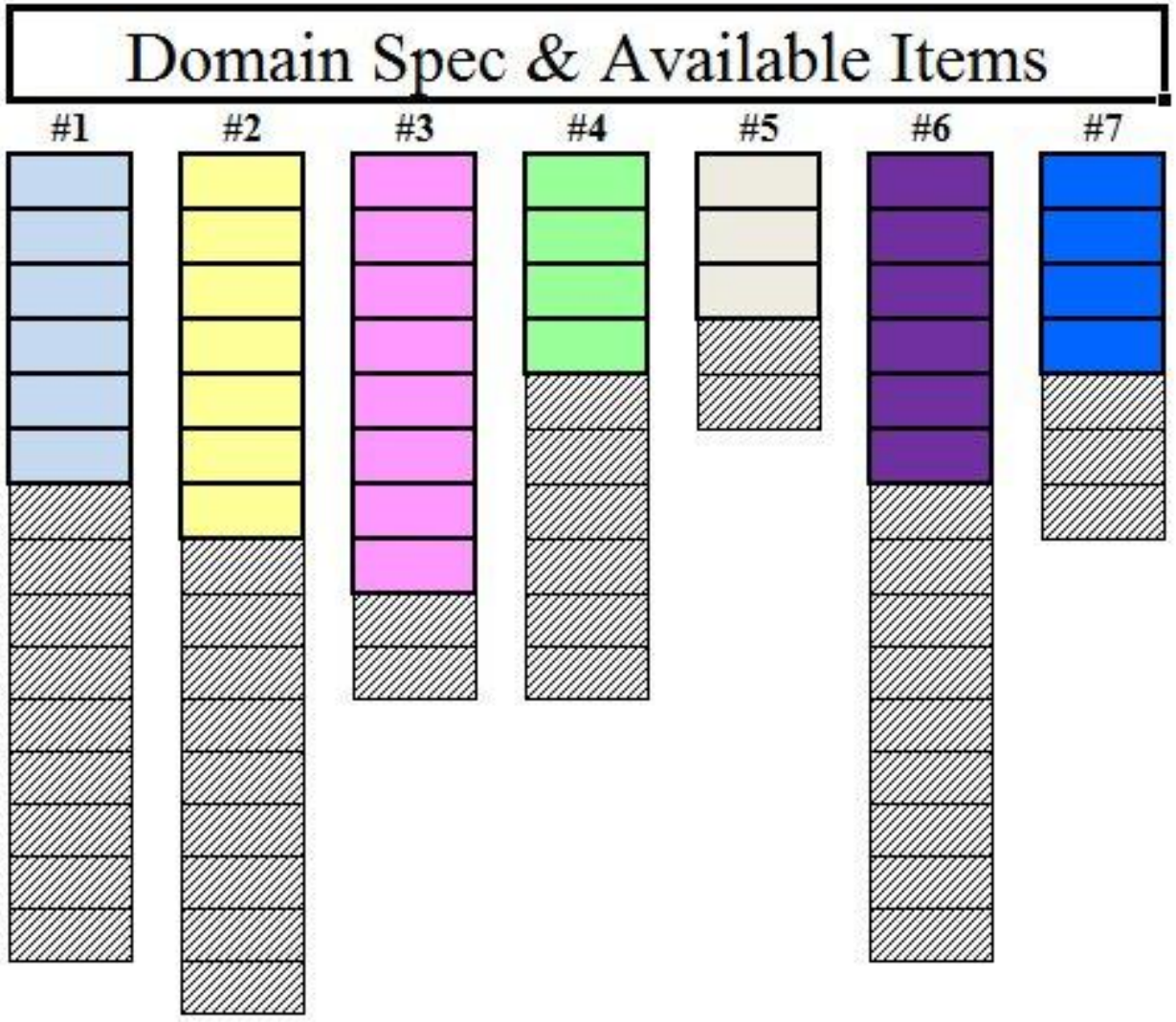
Randomize item seeding for

initial 10 items… from small domains.

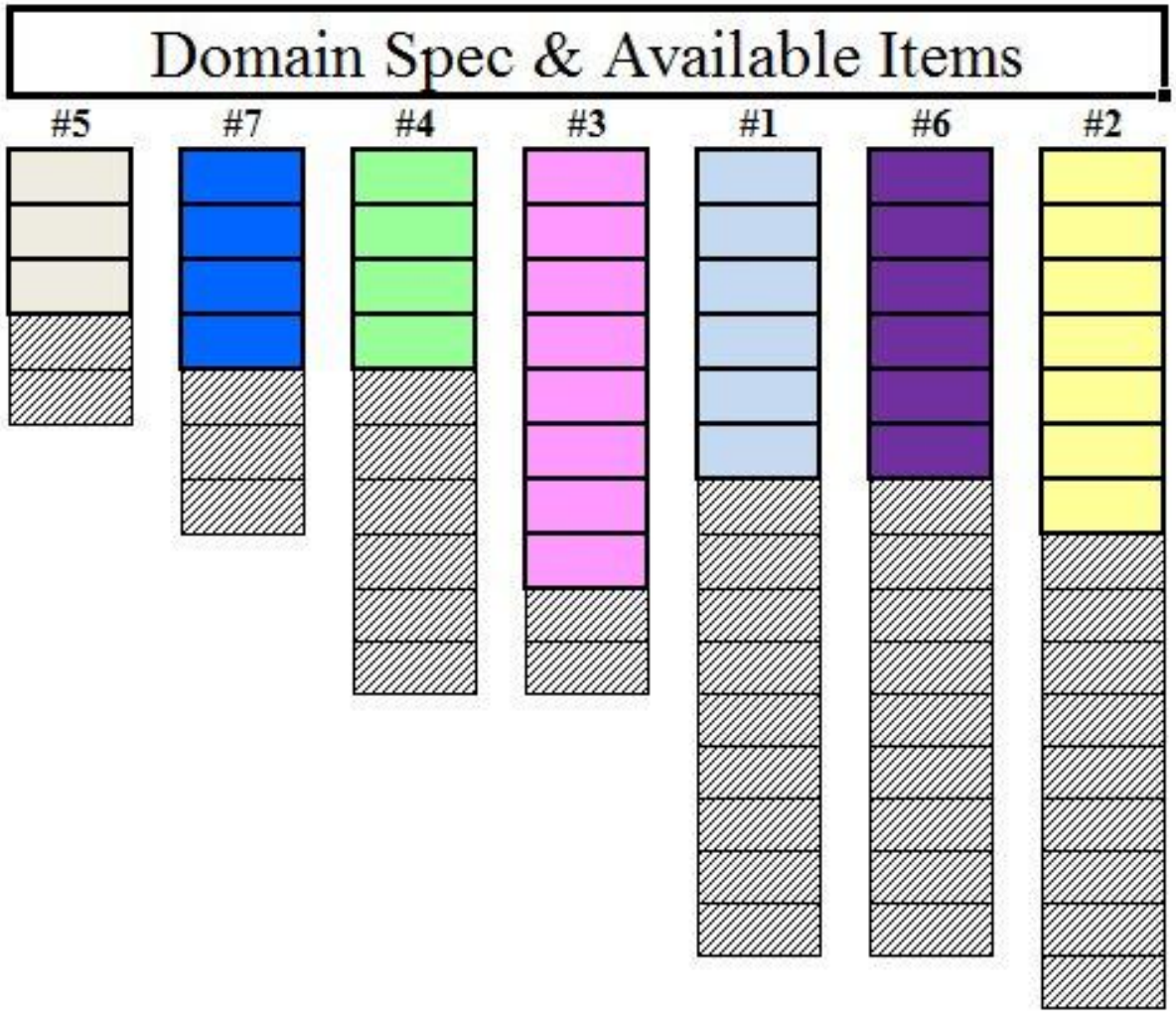
Test Construction Answers:



Test Construction Answers:



Test Construction Answers:



Test Construction

Constraint #2:

Minimize: μ P-Val – Target P-Val

Acceptable: μ P-Val – Target P-Val <= 0.04

Pseudo-Information function drawn from

Classical Test Theory statistics

pInfo (pseudo-Information)

Classical Test Theory statistics:

$P\text{-}Val_i$ = Probability Correct Response, for Item i

$PBis_i$ = Point-Biserial, Item i

$pInfo_i = PBis_i + 1 - (ABS [Cutpoint - P\text{-}Val_i])$

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 Σ 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
- μ P-Val – Target P-Val < 0.04

Test Construction Results

- Domain count consistent with Blueprint
- μ P-Val – Target P-Val < 0.04
- S Latency – Target Response Time < 5.0 min.

Test Construction Results

- Domain count consistent with Blueprint
- μ 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
- μ 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 Σ 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