Comparing the Pros and Cons of Two ATA Methods for Delivering High Volume Pre-Employment Assessments

Li-Ann Kuan, Ph.D., Vice President, Assessment Services

Garrett Sherry, Vice President, Europe Middle East & Africa
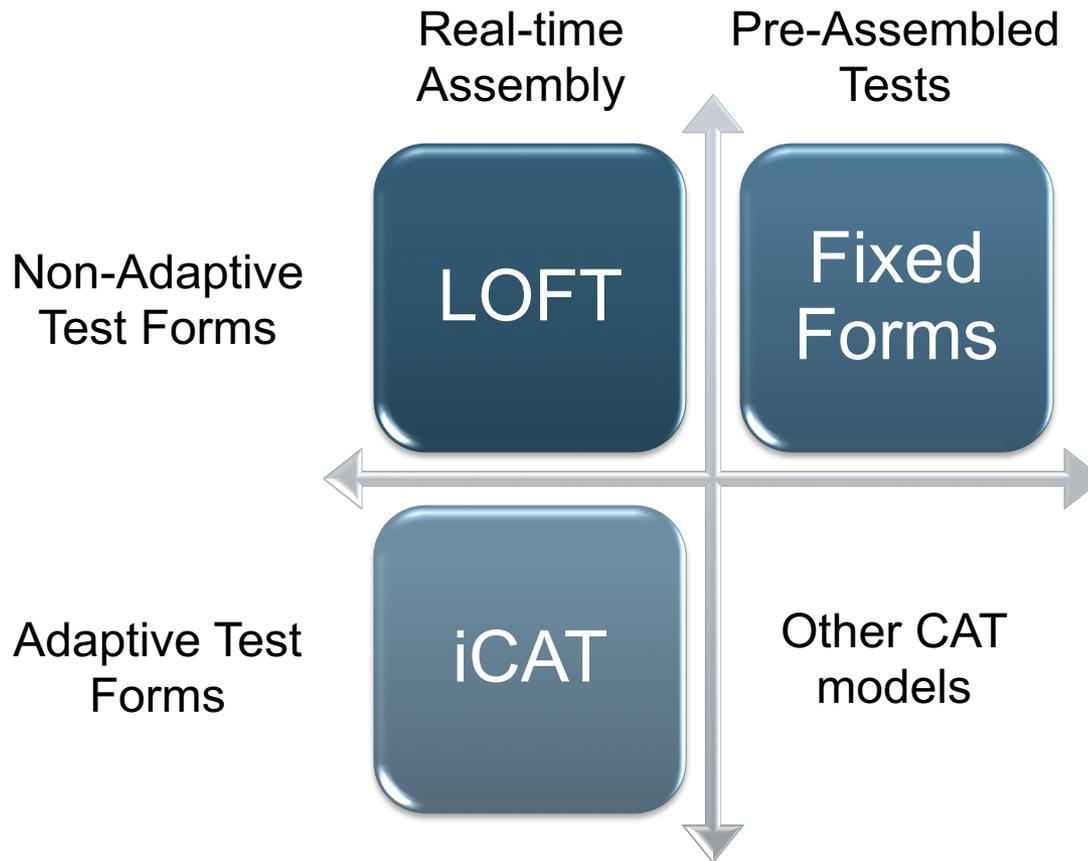
Monday, September 9, 2019

Agenda

+ Prometric Background & Experience with the Public Sector

+ Introduction to Computer Adaptive Tests (CAT) and Linear-on-the-Fly Tests (LOFT)

+ Benefits of CAT and LOFT

+ Things to consider what deciding to use CAT or LOFT

About Prometric

+ HQ Baltimore, USA, EMEA HQ, Ireland

+ Over 300 Clients, Delivering 7M+ tests Annually in 180+ Countries

+ Over 30 Years serving clients in the development & delivery of Computer Based Assessments

+ Working with Public Sector Clients in US, UK, Ireland, Europe & Asia

+ EPSO 'ERICA' Algorithm introduced in 2012 (Modified LOFT)

+ Adaptive Tests not an option for EPSO exams as all forms have to be build according to consistent specifications – same number of items, similar form difficulty

Classes of Test Assembly Methods

	Real-time Assembly	Pre-Assembled Tests
Non-Adaptive Test Forms	**LOFT**	**Fixed Forms**
Adaptive Test Forms	**iCAT**	Other CAT models

Courtesy Richard Luecht

4

A Note About Fixed Forms

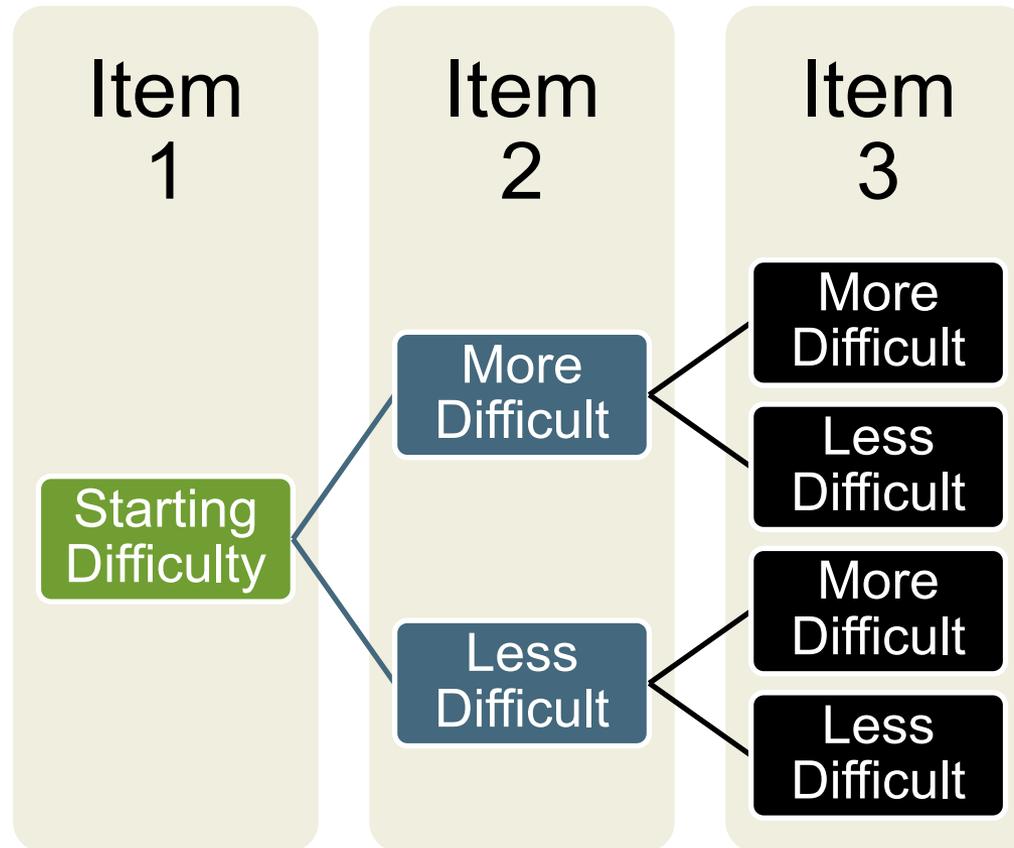+ Fixed-length, linear test forms; employed by a majority of test sponsors

+ A fixed form typically includes:
    - A selection of items that meet content and psychometric requirements
    - A set of common items that allow for test score equating
    - Minimal overlap with other test forms
    - A set of pretest items (optional)

+ Preassembled ahead of the test administration date, allows for quality-control checks (e.g., enemy items, item exposure)

What are Computer Adaptive Tests (iCAT)

+ Test forms that are generated *dynamically* according to predetermined content and psychometric requirements

+ Test is built *in real time* and tailored to each examinee's ability level
  - Each subsequent item is selected based on examinee performance on the previous item
  - Item must closely reflect examinee's ability level, as well as meet content and psychometric requirements

+ Item selection is controlled by a mechanism called *item exposure control* which prevents the overuse of optimal items

+ Item selection stops when:
  - examinee's ability level can be reliably determined
  - the maximum number of items are administered
  - allocated test time expires

6

How CAT Works – Item Selection

Item 1

Starting Difficulty

Item 2

More Difficult

Less Difficult

Item 3

More Difficult

Less Difficult

More Difficult

Less Difficult

How CAT Works – Item Difficulty

+ Each time a candidate answers:

  • incorrectly (0), the difficulty of the next item drops

 • correctly (1), the difficulty of the next item increases

Linear-on-the Fly Tests (LOFT)

PROMETRIC
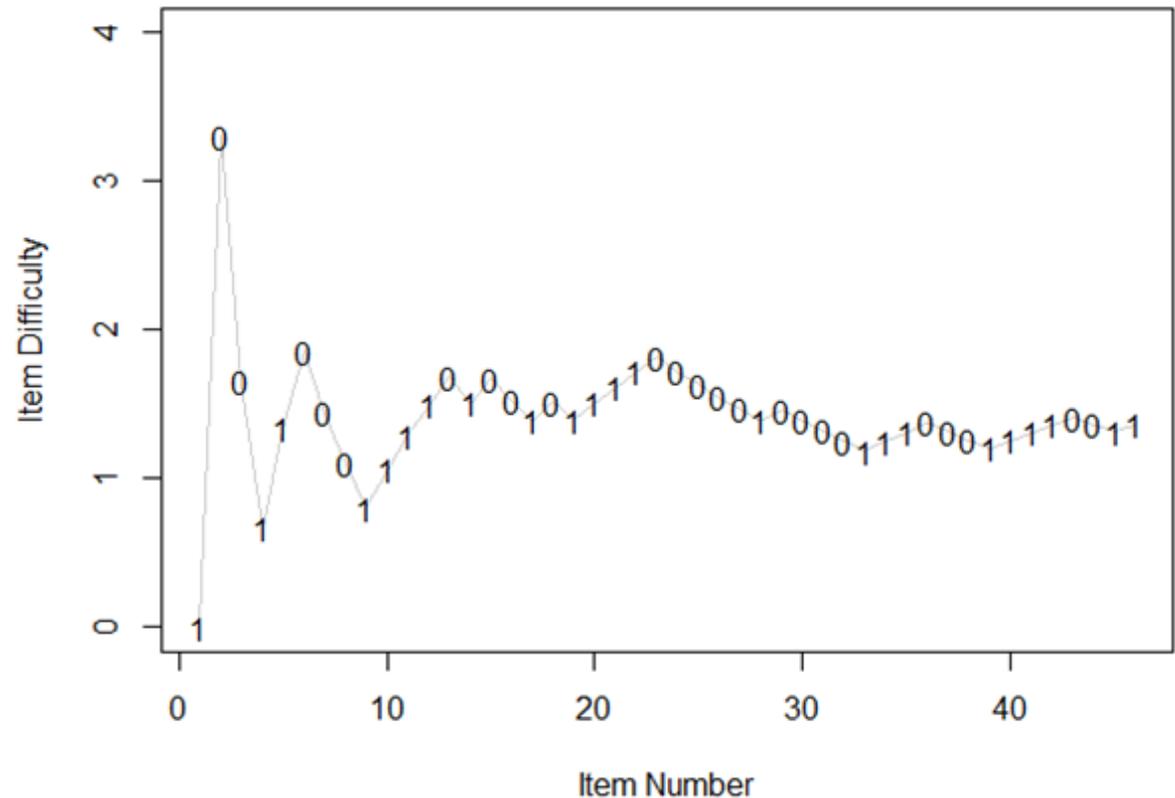
+ Fixed test forms that are _automatically_ generated _in real time_, shortly prior to the testing event

+ Items that meet targeted content and statistical requirements are randomly selected from an item bank

+ Each examinee attempts a _unique test form_ with minimal overlap with other forms

+ Form overlap is controlled by a mechanism called _item exposure control_

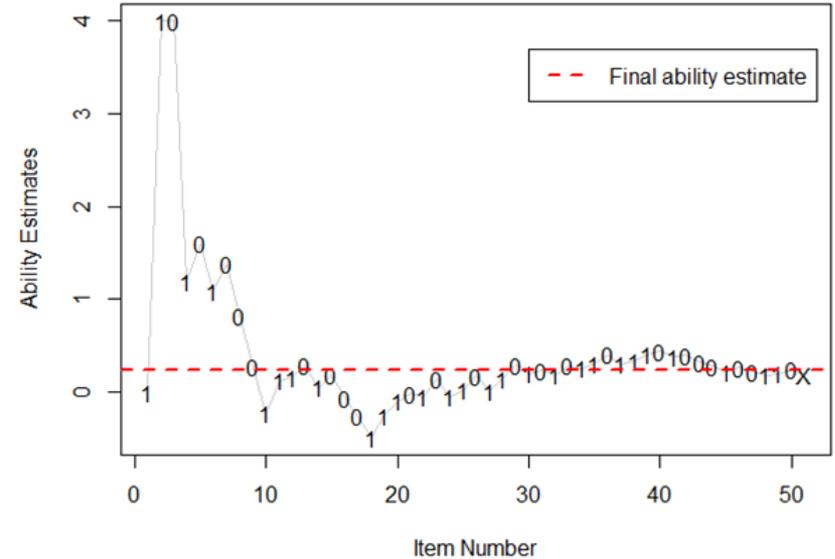# How LOFT Works – Item Difficulty

+ Candidate's successful or unsuccessful response does not determine the difficulty of the subsequent item.
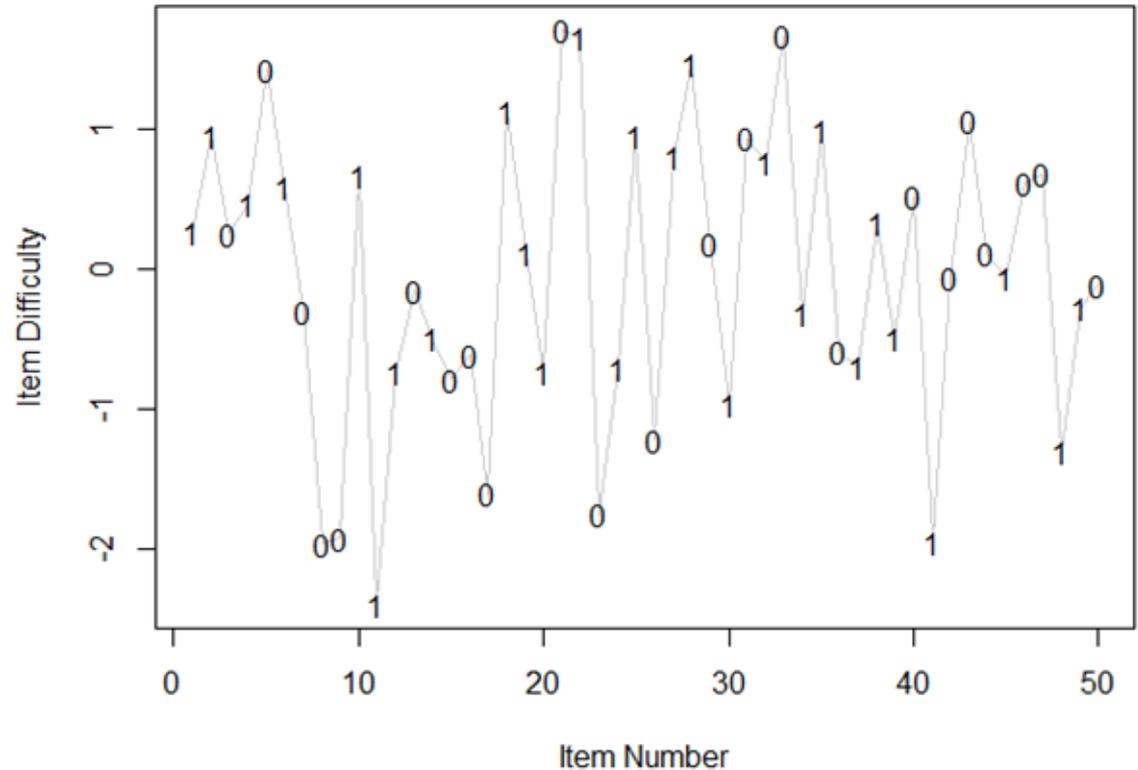
Comparable Ability Estimation
(For Different Exams)
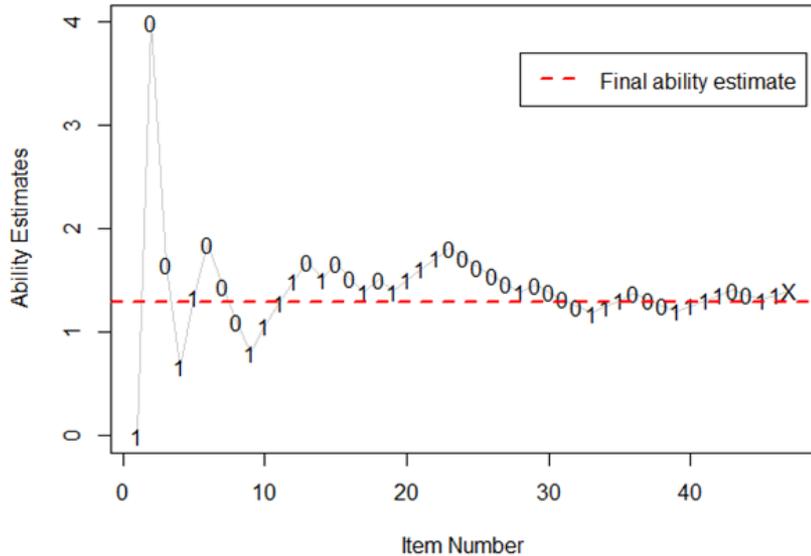
CAT Ability Estimates

LOFT Ability Estimates

+ The ability estimates derived from both CAT and LOFT demonstrate a similar pattern despite variation in the item difficulty

BENEFITS OF CAT AND LOFT

Benefits of CAT and LOFT Over Fixed Forms

+ **Improved security and decreased opportunities for cheating.**

  - unlikely that two examinees receive same set of items

+ **Better usage of all exam items in the bank.**

  - item statistics are kept current

+ **Automated test assembly.**

  - a reduction in error due to manual form assembly

+ **Low form-to-form overlap rates.**

  - low overlap between forms, challenging for examinees to anticipate questions on exam

Unique Benefits of…

CAT

+ **Tests customized to individual ability estimates.**
  - present examinees with items that are likely to provide more information about their ability, increasing measurement precision

+ **Shorter testing time depending on stopping rule.**
  - tailored to an examinee's estimated ability, more information is gained from each item, allows for shorter tests.

LOFT

+ **Consistent test experience for all examinees.**
  - all exam forms are built to the same test specifications (e.g. test length), maintain a similar difficulty and passing standard

+ **Model easier to explain to stakeholders**

THINGS TO CONSIDER WHEN USING CAT OR LOFT

Considerations for CAT and LOFT

+ **Size of the testing program.**
  - rely on IRT methods for initial calibration of items (and item selection during testing for CAT)

+ **Size of the item bank.**
  - large banks required to avoid overexposure of items, ensure that intended test specifications can be adequately covered

+ **Quality of the item bank.**
  - All enemy item relationships must be identified
  - All items must be anchor calibrated using IRT
  - All items must be catalogued to content specifications
  - All items are up to date over time

CAT Considerations

+ **Size of the item bank.**
  - a larger number of items (>10x) required to generate ability estimates

+ **Determining a test stopping rule.**
  - end after specific test length, time limit, or at desired measurement precision

+ **Reliability of real-time internet connectivity.**
  - item bank must be located in the test center driver during administration

+ **Test-taking process.**
  - examinees not allowed to return to items and review or change answers

+ **Publicity and marketing.**
  - clear communication about CAT adoption – candidates may perceive testing experiences as unfair due to variable time or item stopping rules

Prometric's Perspective

+ CAT and LOFT are two different methods of delivering tests – one is not better than the other from a technical standpoint

+ Given profiles of our clients:
  - Complicated test assembly requirements (i.e., enemy items, content and psychometric constraints)
  - Concerns with test security, disinclination to download entire test bank at test delivery sites
  - Limited resources to develop extremely large item banks
  - Desire to maintain high returns on investment, use all items in the bank
  - Desire to keep test specifications consistent; no need to keep forms short due to age of the testing population
  - Desire to allow examinees to change previous responses to items

+ LOFT is a better option than CAT for Prometric's certification/licensure clients

CAT vs LOFT Simulations: Bank Usage

PROMETRIC

Usage Rate	Items (%)		
	LOFT	CAT	
	Precision @ 0.1	Precision @ <= 0.3	Precision @ <= 0.2
65%			<1%
60%			<1%
55%			<1%
50%			1%
45%		<1%	2%
40%			1%
35%		<1%	2%
30%		1%	4%
25%	4%	2%	6%
20%	2%	3%	6%
15%	21%	3%	11%
10%	44%	10%	14%
5-<10%	15%	23%	32%
1-<5%	9%	14%	56%	58%	19%	20%
<1%	2%		2%		1%	
0%	3%		0%		0%	
Total	754	754	754
Avg. Usage	12%	6%	14%

Case Study

+ Higher standard errors of measurement for CAT forms

+ Higher item over exposure rates for CAT items

+ Higher item under exposure rates for CAT

+ In order to support CAT, client would have to increase the number of items in the bank.

Copyright © 2019 Prometric. All rights reserved.

19

CAT vs LOFT Simulations: Test Length

+ All CAT forms shorter than the LOFT forms
  - Majority under 50 items

+ Greater standard error of measurement in CAT forms, i.e., 0.1 vs. 0.3

PROMETRIC

Test Length (# of Test Items)	Number of Test Forms	
	LOFT	CAT
	Precision @ 0.1	Precision @ <= 0.3
46		136
47		266
48		177
49		75
50		40
51		28
52		16
53		6
54		6
55		4
56		2
57		2
…		
64-81		5
90	763	

CAT vs LOFT Simulations: Test Length

+ All CAT forms longer than the LOFT forms

+ Greater standard error of measurement in CAT forms, i.e., 0.1 vs. 0.2

Test Length (# of Test Items)	Number of Test Forms	
	LOFT	CAT
	Precision @ 0.1	Precision @ <= 0.2
90	763	
…		
101		2
102		92
103		193
104		164
105		99
106		66
107		48
108		28
109		14
110-114		23
115-265		34

Questions