

Developing and Publishing a CAT Online

Nathan Thompson

Assessment Systems Corporation

So you want to develop a CAT?

- I receive numerous inquiries every year
- Very few understand what is involved
 - High school teacher with 30 students
 - Professor demanded comprehensive consulting and software for free
- So... what is needed?
 - Good for IACAT to make known

Two main pieces

- Psychometrics and test development
 - Simulation studies
 - Content development
 - Pilot testing and calibration
 - Direction
- Software
 - Develop
 - Purchase/rent/download

An integrated team

- Psychometrician
- Content developer
- Software developer
- Hosting and IT security
- Project manager



1. Psychometrics and TD

Seq.	Stage	Primary work
1	Feasibility, applicability, and planning studies	Monte carlo simulation; business case evaluation
2	Develop item bank content or utilize existing bank	Item writing and review
3	Pretest and calibrate item bank	Pretesting; item analysis
4	Determine specifications for final CAT	Post-hoc or hybrid simulations
5	Publish live CAT	Publishing and distribution; software development

Major issues: Psy and TD

- Most important for psychometrician: Determine CAT is even appropriate or optimal
 - Items scored in real time, format scorable with IRT (or other model)
 - Sample size *and* population size
 - Resources to develop large item banks
 - Adequate staffing and software
 - Examinees need detailed feedback?
 - Any PR issues with a short test?
 - Multidimensionality
 - Bimodal/odd pop. distribution
 - Any other red flags? (you?)

Major issues: Psy and TD

- Determine feasibility, business case
- Ensure adequate item bank
 - Do not over- or under-engineer bank
 - Quality items
- Publish with valid CAT parameters
 - Item selection (exposure, content)
 - Termination
 - Min/max items, time



Major issues: Psy and TD

- Some specific psychometric questions:
 - Bank SEM vs. target SEM
 - Test length and time: 95% approach
 - Exposure constraints?
 - Appropriate min/max
 - Linking design and old/new scale

2. Software

- We've developed a number of tools to work with all these issues
 - Iteman
 - Xcalibre
 - CATSim
 - FastTest Web



2. Software

- Simulation software
 - CATSim (Weiss)
 - FireStar (Choi)
 - WinGen (Han)
- Monte carlo simulations
 - Initial design and planning
- Post hoc and hybrid simulations
 - Establishing valid specifications

2. Software: Running simulations

- 1. Identify IVs:
 - Bank size
 - Bank quality
 - Termination criteria
 - Constraints
- 2. Identify DVs
 - Test length
 - CAT accuracy (PCC, correlation with true)
- 3. Run
- 4. Collate and evaluate

2. Software

- CATSim

Initial Theta

- Initial theta for all examinees is
- Initial theta is random in the interval to
- Read initial thetas and SEMs for each examinee from a file



Use individual SEMs as Bayesian prior SDs when Bayesian scoring is selected

Estimate Theta By

- Maximum likelihood: Estimate theta for initial items for non-mixed response patterns using
 - A step size of theta/difficulty =
 - Bayesian estimation with a mean of and a standard deviation of
- Bayesian estimation with a mean of and a standard deviation of
For Bayesian estimation, estimate theta by EAP (expected a posteriori) MAP (maximum a posteriori)
- Weighted maximum likelihood

Item Selection Option

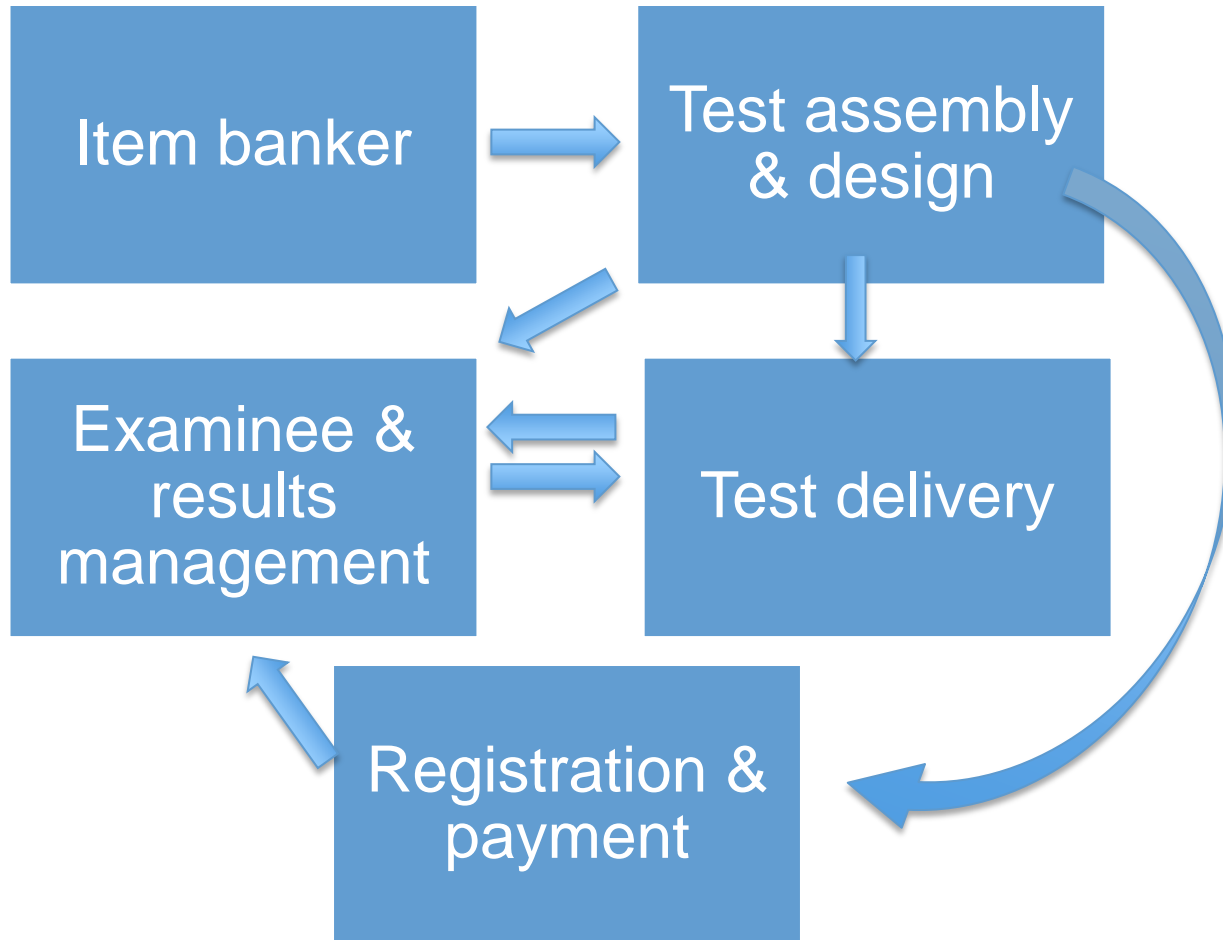
- Select all items by maximum information at the current theta estimate
- Randomly from the items with maximum item information for the first items in the test
- Select items in order of maximum information at theta =

2. Software: **FastTEST Web** Innovative solutions for online testing

- Goal: System to manage entire process
 - Item bank development
 - Pilot testing
 - Psychometric analysis (IRT in Xcalibre)
 - CAT design
 - Registration/payment
 - CAT delivery
 - Results management



2. Software: Modules



2. Software: Key issues

- Professional item banker
- Multiple item types
- Workflow and quality control
- Point and click design (ASC goal!)
- Advanced models (polytomous, CCT...)
- Comprehensive, enterprise-class
- Generally usable yet customizable
- Tablets

2. Software: Key issues (future)

- Non-admin portals (student, teacher...)
- Integrated test batteries
- Advanced psychometrics
- Interactive technology
- Remote proctoring
- Standards (QTI, APIP)
- International hosting

2. Software: Key issues

- **Security!!!**
 - Hosting in professional data center with automatic offsite backup, load balancing...
 - Firewalls and protocols
 - System logs
 - N-tier structure
 - User roles and area limits
 - WebLock and test passwords



Functionality to consider

- Item banking
 - (see next page)
- Test form assembly
 - Preview
 - Print or export to 3rd party
 - Publish online
- Delivery
 - On-demand computerized
 - Linear
 - Randomized
 - CAT/CCT
 - Traditional paper-and-pencil
 - Configurable security (WebLock)
- Registration
 - Examinee, proctor, test centers, e-commerce, API
- Operations Management
- Reporting
 - Examinee feedback (diagnostics)
 - Examinees/groups
 - Item banks
 - Classical analysis
 - IRT analysis



Functionality to consider

- Item banking
 - Item authoring & importing
 - Item review
 - Standard setting
 - Metadata/metatags
 - IRT and CTT statistics
 - User-defined fields
 - Multimedia
 - Versioning and history
 - Rationale
 - Learning objectives



Functionality to consider

- System management
 - Hierarchy of content within an account (workspace)
 - Multiple workspaces
 - “Publishing” to your clients (subworkspaces)
 - Users and roles
 - Locations
 - Accounting
 - Time expirations
 - Custom/specific stuff
 - Confirmation testing for pre-employment tests
 - Tracking K-12 scores (retesting)

Educating stakeholders

- What can CAT do, what it can't
- What scores mean
- What is the purpose of testing? To produce accurate scores for all examinees, which CAT does and fixed form does not
- How CAT works
- What examinees will see
- How it increases security and other benefits

2. Software

- Show FastTest Web?
- Slide here on educating stakeholders!
- Thank you! nthompson@assess.com