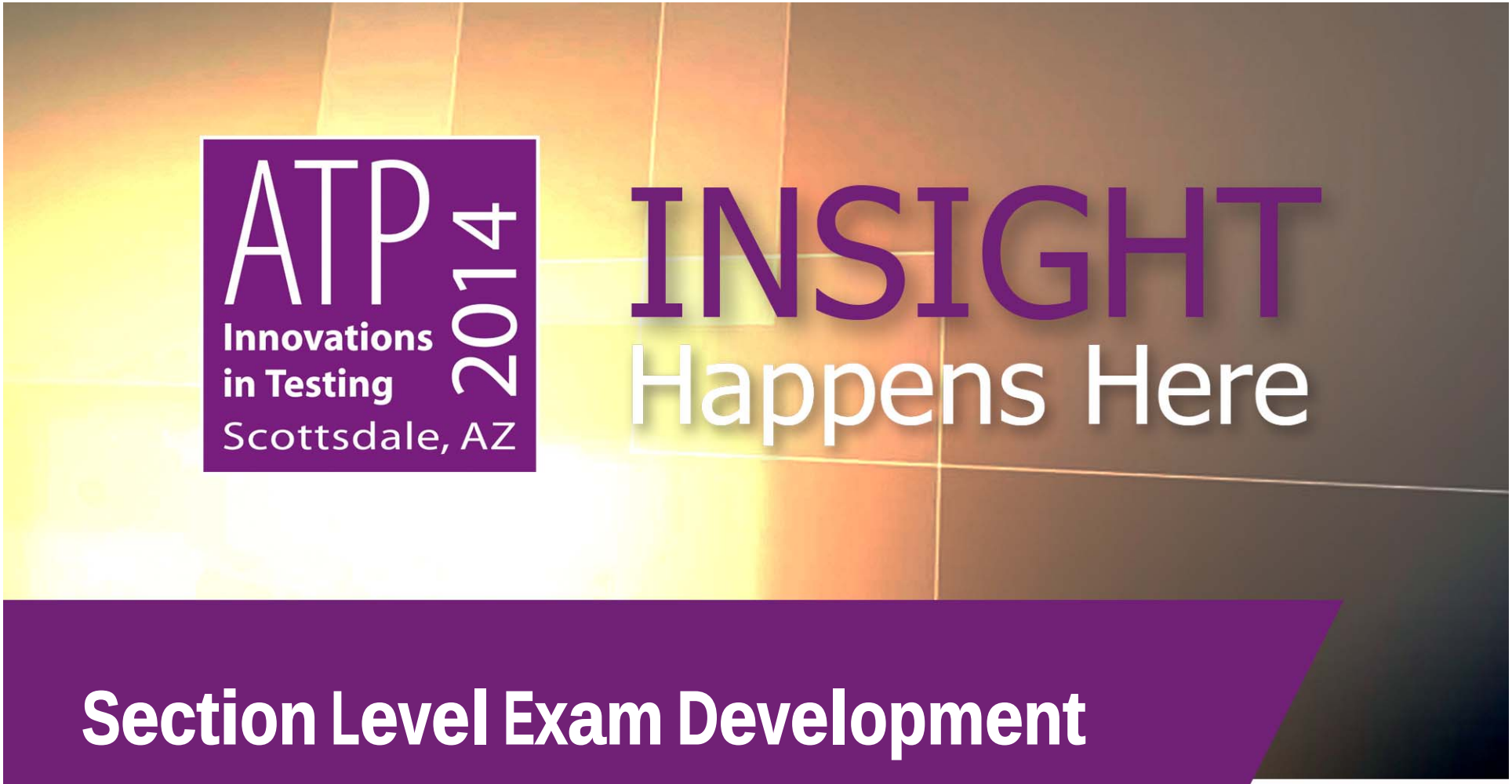


Section Level Exam Development

In Response to Rapidly Changing Content Domains

Jessi Mielke, Esri & Lisa O'Leary, Ph.D., Alpine Testing Solutions

Introduction & Agenda

- What is Section Level development?
  - Development of exams comprised of section-level testlets with distinct content and statistical properties
- Program considerations
- Psychometric & exam development considerations

Assessing Program Need

- How do you determine there is a need to move to section-level development?
- Contributing factors
  - Organizational change
  - Candidate needs
  - Program health
  - Other considerations

Preparing for Section-Level Development

- How do you prepare for section level development?
  - Clear information regarding changes
  - Input from appropriate players

Exam Development

- Preparing your schedule
  - Plan ahead
  - Increase in workshop frequency
    - New workshop types
  - SME engagement
  - Update rules
    - Level of changes (% threshold for changes)
    - Time, health, future releases
  - Other considerations

Exam Development

- Adjust approaches to development
  - Consistent review of exam content
  - Introduce new techniques
  - Working with SMEs
  - Maximize productivity

Exam Development

- Section-level development affects processes and deliverables throughout the exam development cycle

Exam Development

- Section-level development might necessitate order changes to typical exam development schedule

Exam-Level

Design Program ➡ Design Test ➡ Analyze Domain ➡ Develop Blueprint ➡ Develop & Review Content ➡ Pre-Test & Analyze ➡ **Assemble Operational Test** ➡ Conduct Standard Setting ➡ Maintain Test

Design Program ➡ Design Test ➡ Analyze Domain ➡ Develop Blueprint ➡ Develop & Review Content ➡ Pre-Test & Analyze ➡ Assemble Operational Test ➡ **Conduct Standard Setting** ➡ Maintain Test

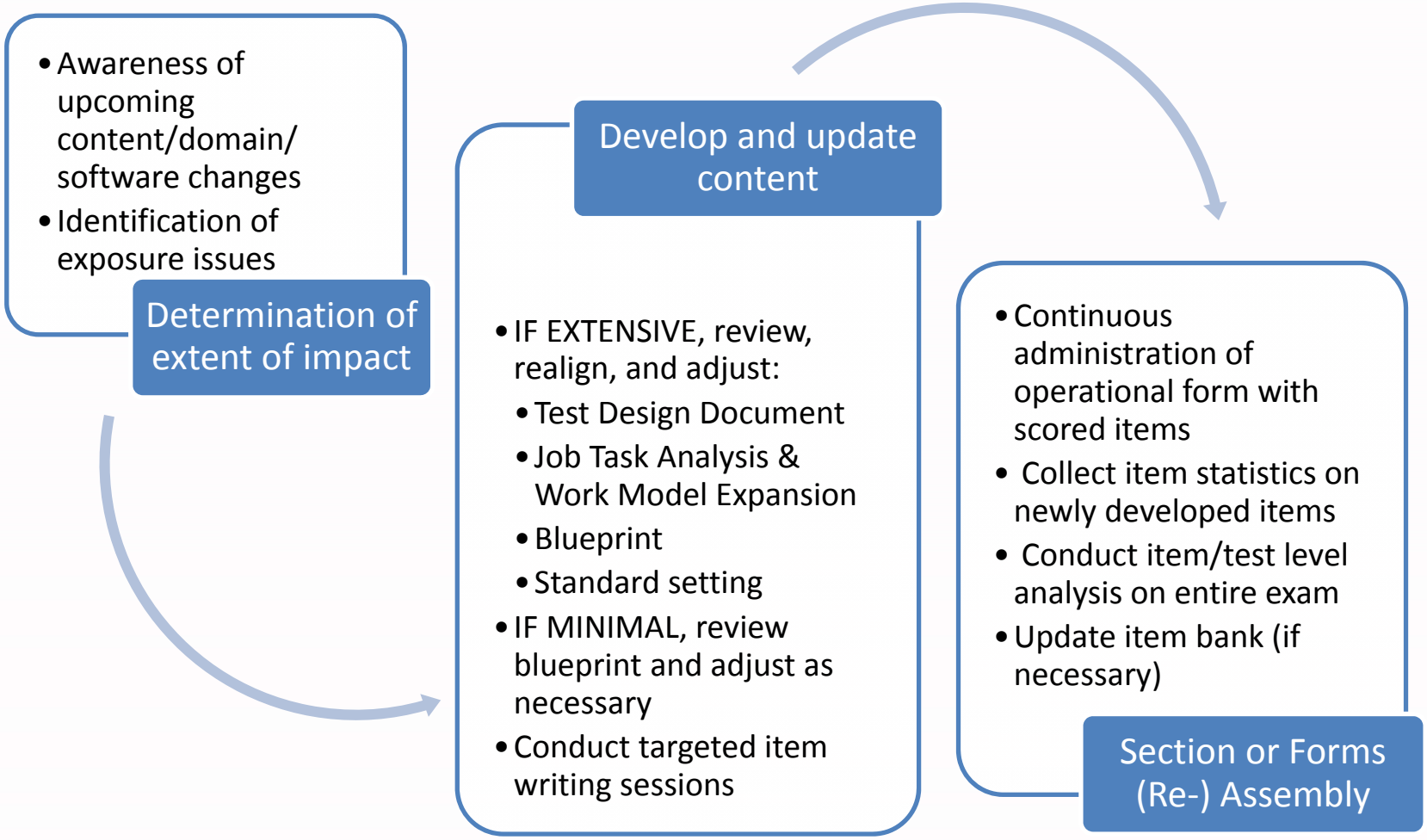
Section-Level

Design Program ➡ Design Test ➡ Analyze Domain ➡ Develop Blueprint ➡ Develop & Review Content ➡ Pre-Test & Analyze ➡ **Conduct Standard Setting** ➡ Assemble Operational Test ➡ Maintain Test

Design Program ➡ Design Test ➡ Analyze Domain ➡ Develop Blueprint ➡ Develop & Review Content ➡ Pre-Test & Analyze ➡ Conduct Standard Setting ➡ **Assemble Operational Test** ➡ Maintain Test

Exam Maintenance

- Iterative cycle based on frequency and extensiveness of changes



Exam Blueprint Structure

- Identify and categorize the impact of upcoming changes that go beyond the skills expected of the current Minimally Qualified Candidate (MQC)

Impact Categories

Rank	Impact Category	Description
0	No impact	· Objective **not** impacted by 10.2. · All items are very likely to be approved, as is, for 10.2.
1	Minimal	· Objective **minimally** impacted by 10.2 - Consideration to minor improvements or cosmetic · All items are likely to be approved for 10.2, with some items needing simple edits.
2	Minor	· Objective **somewhat** impacted by 10.2 - Minor functionalities, tools, or enhancements may have been introduced, but 10.1 workflows mostly remain the same at 10.2. · Many items are likely to be approved at 10.2, but some may require minor revision effort on behalf of a SME.
3	Moderate	· Objective is **moderately** impacted by 10.2 - Major functionalities, tools, or enhancements have been introduced. Some features have been deprecated or no longer considered a best · Impact will vary across items, as some may be approved, reworked, or rejected. New items will need to be written to test knowledge on new functionality.
4	Major	· Objective is **significantly** impacted by 10.2 - New, major workflows are introduced or have been dramatically improved. · Almost all items are likely to be rejected. New item content for this objective is a high priority for this exam.
5	Obsolete	· Objective **no longer applies** at the 10.2 version due to radically different functionality. · We expect the objective to be replaced entirely or existing content may be rolled into another objective.

Exam Blueprint Structure

- Make initial determinations of possible blueprint restructuring to better align with updated content

Esri 10.1 - Blueprint		SME1	SME2	Category: Major	
Section 1	**Product knowledge**	**Impact**	**Impact**	**Describe 10.2 impact**	**Additional Notes**
Objective 1.1	Explain fundamental (as it relates to enterprise system design) database concepts (e.g., replication, versioning, ArcSDE-supported spatial data types, database joins, normalization, editor tracking, direct a database connection, etc.).	2	3	SME1 - Increased capabilities from native database tables (editable feature services hosted directly by ArcGIS Server) SME2 - AGS 10.2 can publish editable feature services from supported DBMS - new design solutions to integrate with non-geodatabase attribute tables.	
Objective 1.2	Compare and contrast the different caching options (e.g., type of caching, cache on demand, partial cache, full cache, output formats, fused or multi-layer, exploded, compact, etc.).	1	3	SME1 - at 10.1 SP1, Importing/Exporting caches have the option to overwrite destination cache SME2 - New cachingcontrollers service manages all jobs processed by the CachingTools service - controls maximum number of CachingTools instances.	
Objective 1.3	Compare and contrast the dynamic vs. cached datasources (e.g., operation layers, base map layers, level of effort to build/update, maintenance,	0	0		SME1 - I'd probably merge with 1.2 SME2 - this could be merged with 1.2
Objective 1.4	Describe the functional capabilities of each service type (e.g., map service, geoprocessing, image services, etc.).	1	0	SME1 - Minor tweak to Map, Feature, image, WFS service - uses standardized queries SME2 - Minor improvements not significant for MQC.	SME2 - This could be merged with 1.5
Objective 1.5	Describe the extended capabilities of extension service types (e.g., network analyst, spatial analyst, business analyst, etc.)	3	3	SME1 - New Extensions: GeoEvent Processor SME2 - Geoevent processor, real time displays, big data - these extend the solution options and are important concepts for MQC. Development integration efforts with IBM Hadoop environments provide more options for collection and query of big data sources.	SME1 - merge with 1.4 - a service is a service is a service, the distinction between this point and 1.4 is a licensing question, not identification

Exam Blueprint Structure

- Conduct blueprint review meeting with SMEs to restructure the blueprint for section-level content updates
  - Confirm SME suggestions in a follow-up meeting with a blueprint survey

10.1 Blueprint

Section	Content	Blueprint	Target BP
1	Product Knowledge	25.3%	24
2	Performance and scalability	6.3%	6
3	Design considerations	4.2%	4
4	Identify the business goals, outputs, and processes.	2.1%	2
5	Conduct a user needs assessment and define the use cases.	3.2%	3
6	Gather non-functional requirements and service level requirements.	1.1%	1
7	Map the business requirements to the technology.	5.3%	5
8	Design the conceptual architecture	3.2%	3
9	Design the business architecture	1.1%	1
10	Design the application architecture	9.5%	9
11	Design the data architecture	8.4%	8
12	Design approach to security	6.3%	6
13	Design licensing	3.2%	3
14	Design integration and interoperability	6.3%	6
15	Design the deployment architecture	4.2%	4
16	Capacity planning and performance	4.2%	4
17	Documentation	2.1%	2
18	Build phase/migration plan	3.2%	3
19	Solutions development	1.1%	1

10.2 Blueprint

Section	Content	Blueprint	Target BP
1	Capacity Planning, Performance, and Scalability	18.9%	18
2	User Needs Assessment	10.5%	10
3	Technical Architecture (Hardware and Network)	12.6%	12
4	Application Architecture	14.7%	14
5	Data Architecture	17.9%	17
6	Security	10.5%	10
7	Design, Implementation, and Migration Plan	14.7%	14

Target Number of Items

- Write or retain a sufficient number of items to reliably report at the section-level; also translates into high exam-level reliability

10.1 BLUEPRINT

Section	Item Count	Section Reliability	# of Items for Section Reliability ≥ 0.6
1	24	0.608	23
2	6	-0.264	18
3	4	-2.969	18
4	2	-12.636	18
5	3	-3.978	18
6	1	*	18
7	5	-3.432	18
8	3	-4.134	18
9	1	*	18
10	9	-0.374	18
11	8	-0.105	18
12	6	-0.852	18
13	3	-4.961	18
14	6	-1.212	18
15	4	-3.283	18
16	4	-2.208	18
17	2	-30.073	18
18	3	-7.443	18
19	1	*	18

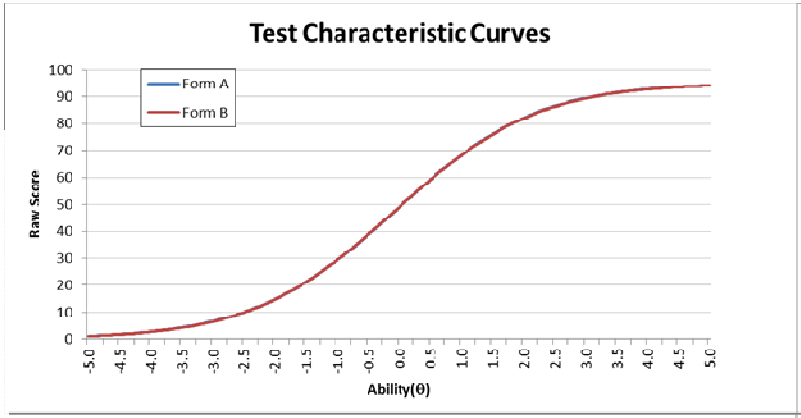
10.2 BLUEPRINT

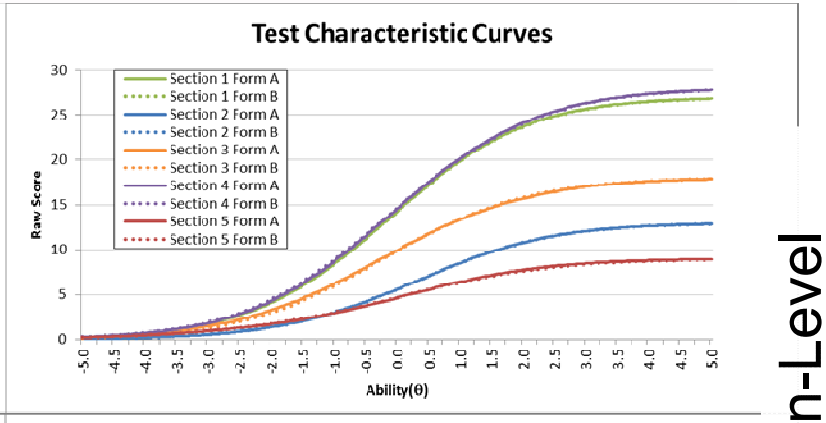
Section	Item Count	Section Reliability (Estimated)	# of Items for Section Reliability ≥ 0.6
1	18	0.605	18
2	10	0.460	18
3	12	0.505	18
4	14	0.544	18
5	17	0.591	18
6	10	0.460	18
7	14	0.544	18

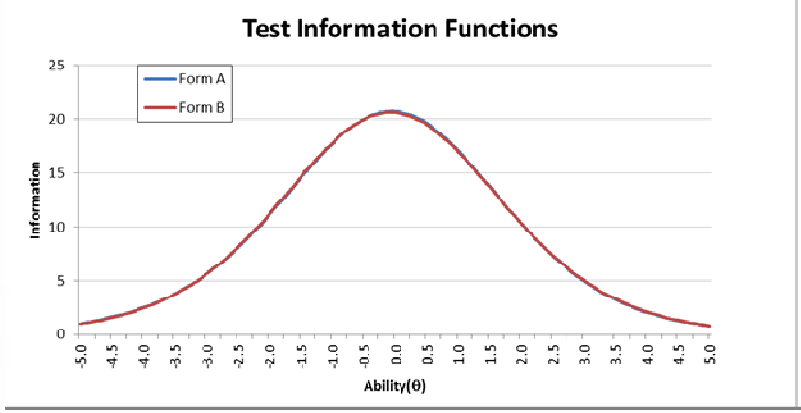
Continuous Form Administration

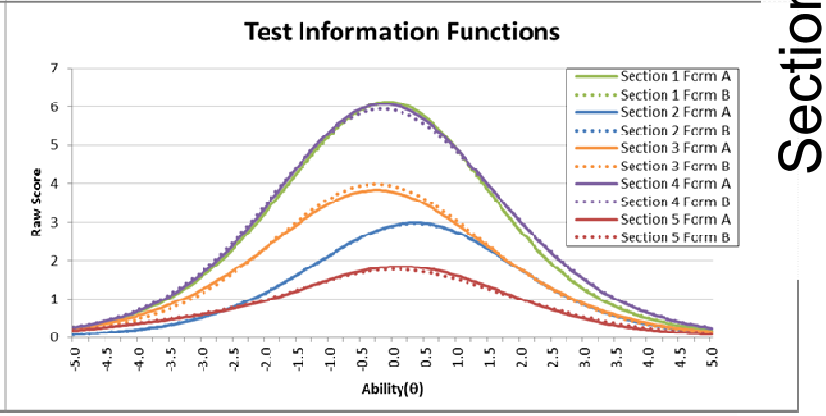
- Pre-equate at the section-level to ensure fair and consistent scoring and comparability of test scores across different forms

Exam-Level









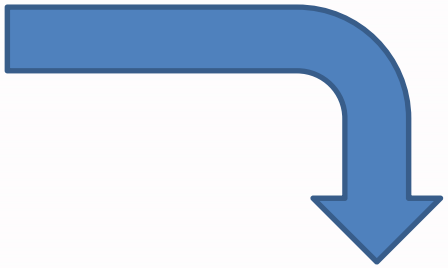
Section-Level

Continuous Form Administration

- Build and balance at the section-level during forms assembly to allow for future section-specific updates

Exam-Level

	Form A	Form B
Number of items	95	95
Mean	56.54	56.66
Standard deviation	16.40	16.54
Alpha Reliability	0.93	0.93
Average Test Time (minutes)	86.20	85.63
Estimated number correct at target cut score	58.42	58.40
Percent correct at target cut score	61.49%	61.47%

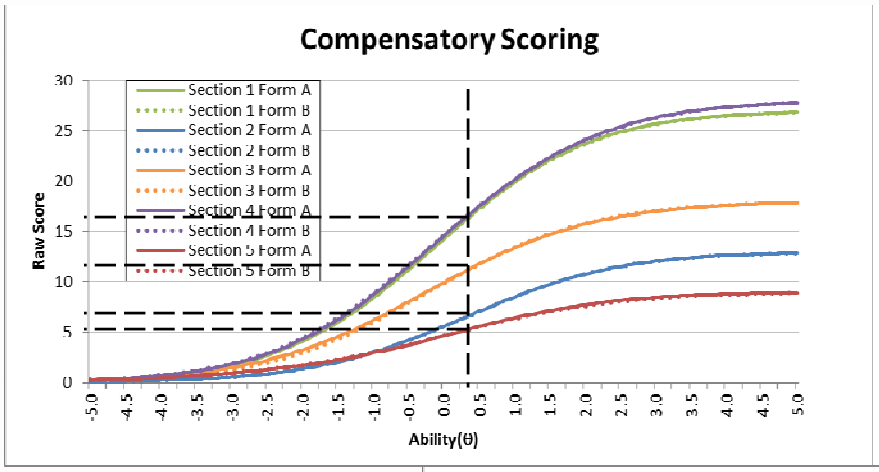


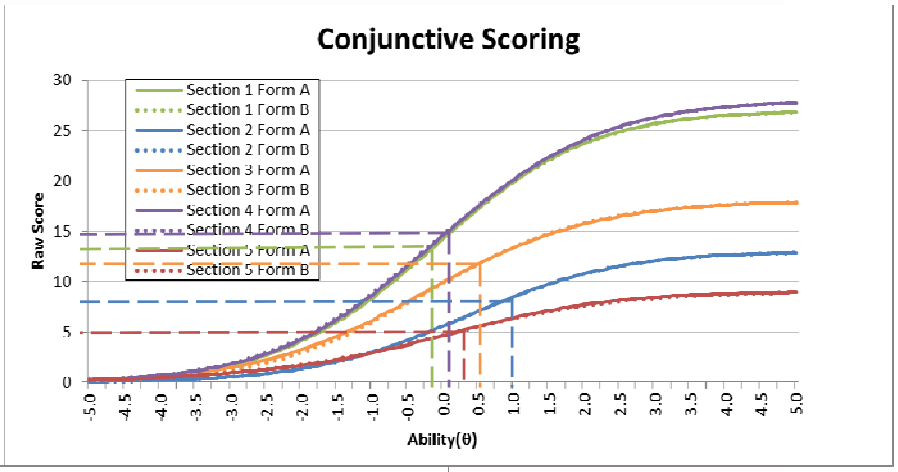
Section-Level

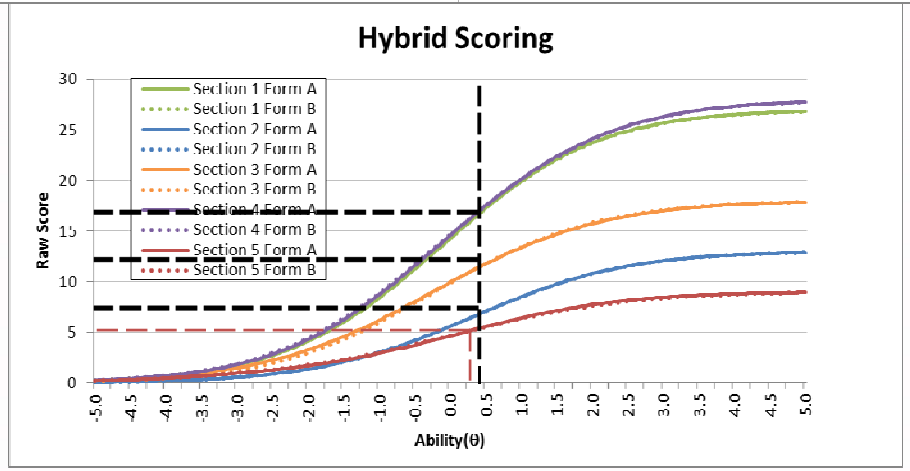
	Section 1		Section 2		Section 3		Section 4		Section 5	
Stats	Form A	Form B	Form A	Form B	Form A	Form B	Form A	Form B	Form A	Form B
Number of items	27	27	13	13	18	18	28	28	9	9
Mean	16.44	16.47	6.79	6.83	11.24	11.22	16.73	16.79	5.34	5.35
Standard deviation	5.36	5.51	2.59	2.68	3.30	3.56	4.85	4.77	1.62	1.34
Alpha Reliability	0.821	0.834	0.594	0.626	0.701	0.739	0.774	0.769	0.328	0.010
Average Test Time (minutes)	26.93	25.99	10.58	11.03	17.43	16.71	24.23	24.52	7.03	7.38
Average Item Measure	-0.13	-0.14	0.30	0.30	-0.24	-0.20	-0.07	-0.08	-0.20	-0.18
Estimated number correct at target cut score	17.03	17.02	6.96	6.95	11.63	11.63	17.30	17.31	5.49	5.48
Percent correct at target cut score	63.06%	63.05%	53.57%	53.50%	64.62%	64.60%	61.79%	61.83%	61.04%	60.89%

Scoring Options

- Compensatory, conjunctive, or hybrid scoring options







Scoring Options – Score Reports

- Present diagnostic feedback to candidates on section-level pass or fail performance based on equated cut scores

Enterprise Administration Associate (EEAA)
Performance Report

The information in the table below details the composition of the EEAA exam and your performance in each of its **5 sections**. The table includes the percentage of the exam that was dedicated to each content area and classifications of your performance at each **section-level**.

Pass: Performance at this level demonstrates that expected of a passing candidate.
Fail: Performance at this level fails to meet that expected of a passing candidate.

Section	Percent of Scored Items	Score Performance Level	
		Fail	Pass
Implement and Deploy a Solution	28.4%	x	
Maintain and Support a Solution	13.7%		x
Troubleshoot Problems with ArcGIS Server	18.9%		x
Prepare and Publish Content	29.5%	x	
Portal for ArcGIS	9.5%	x	

Disclaimer: The EEAA 10.2 exam was designed to make pass/fail decisions at the overall exam-level and allow for a summary of section-level performance. The overall pass/fail designation is a representation of the determination of candidates' knowledge, skills, and abilities at the overall exam-level. The section-level information can be considered diagnostic feedback of performance in particular content areas. Although pass/fail decisions were not made based on candidates' individual section-level scores, candidates can interpret the above section-level score information as a guide for future test preparation.

Certification Program Considerations

- Program design
  - Exam names
  - Versioned exams
  - Exam maintenance
  - Candidate communication
- Candidate policies
  - Exam retakes
  - Recertification

References

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (1999). *Standards for educational and psychological testing.* Washington, DC: AERA.

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Mahwah, NJ; Lawrence Erlbaum Associates, Publishers.

Crocker, L., & Algina, J. (1986). *Introduction of classical and modern test theory.* New York, NY: Harcourt Brace Jovanovich College Publishers.

Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking.* New York, NY: Springer.

Lord F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading, MA: Addison Wesley.

Rasch, G. (1960) *Probabilistic Models for Some Intelligence and Attainment Tests*. (Copenhagen: 1960) Chicago, IL: MESA Press, 1992

Wright, B.D., & Stone, M.H. (1979). *Best Test Design*. Chicago, IL: MESA Press.

Thank you!

Jessi Mielke, Esri Certification Program Manager (jmielke@esri.com)
Lisa O'Leary, Ph.D., Alpine Testing Solutions Psychometrician (lisa.oleary@alpinetesting.com)