

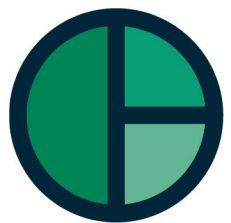
Institute for Credentialing Excellence

EXG2023



OCTOBER
24-27
2023

COLORADO SPRINGS, CO



Institute for Credentialing Excellence

EXG2023

Advancements in job analysis and content outline development

Rob Brucia, PhD, National Association of Boards of Pharmacy

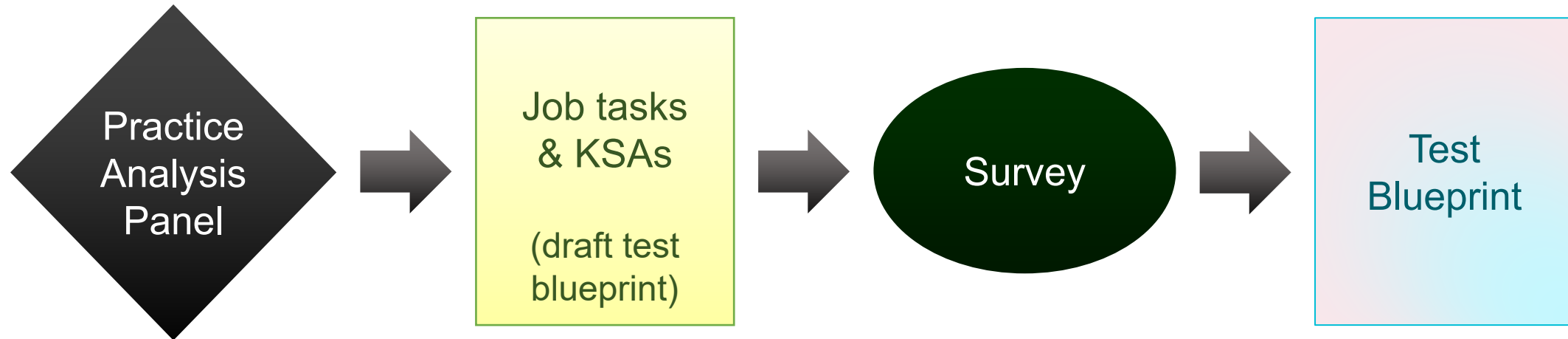
Andy Dwyer, PhD, American Board of Pediatrics

Brett Foley, PhD, Alpine Testing Solutions





Job Analysis - Basic Process



Session Objectives

Present research, case studies, and ideas that may:

- 1) Improve the relevance of the exam (and, by extension, the validity of certification decisions based on exam performance)
- 2) Improve the operational efficiency of the job analysis process

Agenda

1. Rating scale comparison study (Rob Brucia)
2. Content domain vs. item-level relevance ratings study (Rob Brucia)
3. NCARB case study (Brett Foley)
4. ChatGPT demo and discussion (Brett Foley)
5. Test blueprint/content outline guidance (Andy Dwyer)

Survey rating scale study

Common Rating Scales

- Single Scales
 - Criticality
 - Frequency
 - Time spent
 - Importance
 - Point of acquisition
 - Need at entry
 - Difficulty of learning
- Composite Scales
 - $\text{Criticality} * \text{Frequency} = \text{“Relative Importance”}$
 - Importance as a scale can be subjective
 - Some composite scales may weigh one scale higher than another (e.g., $\text{Criticality}(2) + \text{Frequency}$)

ABP Rating Scales

Traditional

- Criticality
- Frequency
- $\text{Importance} = \text{Criticality} * \text{Frequency}$

New

- Relevance

Motivation for study

- Started collecting relevance ratings at the item level through our longitudinal assessment, MOCA-Peds (more on this later)
- Declining response rates to practice analysis surveys

Research Questions

- Are these rating scales redundant or do they provide unique and useful information?
- Would using fewer rating scales reduce survey burden (response time) and/or increase response rates?
- If using multiple rating scales, does presentation order matter?
- **Which rating scale(s) should we use?**

Rating scales

- Criticality
- Frequency
- Importance (Crit * Freq)
- Relevance

Study rating scales

Frequency

“Frequency with which you use knowledge and skills in each content area”

- Never (1)
- Yearly (2)
- Monthly (3)
- Weekly (4)
- Daily (5)

Criticality/Harm

“Level of harm that would be caused by a lack of knowledge and skills in each content area”

- Little or no harm (1)
- Moderate level of harm (2)
- Serious or severe harm (3)

Relevance

“Relevance of knowledge and skills in each content area to your practice”

- Not at all relevant (1)
- Slightly relevant (2)
- Moderately relevant (3)
- Very relevant (4)

“Importance” = Freq * Crit

Study design

- **Pediatric Cardiology**

- N = 2,000
- 176 content areas
- 3 survey versions
 1. Relevance
 2. Frequency
 3. Criticality

- **Pediatric Neonatology**

- N = 5,119
- 136 content areas
- 3 survey versions
 1. Relevance
 2. Frequency
 3. Criticality

- **General Pediatrics**

- N = 77,599
- 225 content areas
- 11 survey versions

Survey Version

- 1 – Relevance
- 2 – Frequency
- 3 – Criticality
- 4 – Relevance + Domain Weights + BMH
- 5 – Relevance + Criticality
- 6 – Criticality + Relevance
- 7 – Criticality + Frequency
- 8 – Frequency + Criticality
- 9 – Frequency + Relevance
- 10 – Relevance + Frequency
- 11 – Open comments + Domain Weights + BMH

Results: Response Rates (Gen Peds)

Survey Version	N Invited	Started survey		Completed survey	
		N	%	N	%
1 – Relevance	9,700	1,222	12.60	783	8.07
2 – Frequency	9,700	1,212	12.49	759	7.82
3 – Criticality	9,700	1,258	12.97	688	7.09
4 – Relevance + Weights + BMH	9,700	1,191	12.28	686	7.07
5 – Relevance + Criticality	4,849	633	13.05	321	6.62
6 – Criticality + Relevance	4,850	622	12.82	288	5.94
7 – Criticality + Frequency	4,850	604	12.45	275	5.67
8 – Frequency + Criticality	4,850	628	12.95	290	5.98
9 – Frequency + Relevance	4,850	583	12.02	291	6.00
10 – Relevance + Frequency	4,850	589	12.14	284	5.86
11 – Remove/open comments + Weights + BMH	9,700	1,238	12.76	634	6.54
Total	77,599	9,780	12.60	5,299	6.83

Results: Correlations

	Frequency & Criticality	Relevance & Frequency	Relevance & Criticality	Relevance & Importance (Freq*Crit)
Pearson Correlation				
General Pediatrics	-0.172	0.901	0.101	0.922
Pediatric Cardiology	0.357	0.620	0.889	0.891
Pediatric Neonatology	0.400	0.767	0.798	0.896
Spearman's Rank-Order Correlation				
General Pediatrics	-0.213	0.896	0.082	0.933
Pediatric Cardiology	0.365	0.626	0.889	0.897
Pediatric Neonatology	0.416	0.763	0.803	0.911

Next steps

- ABP has decided to only use the Relevance rating scale in practice analysis surveys going forward
- While response rates did not appear to be impacted by one versus multiple rating scales, the use of a single relevance scale will minimize survey burden and allow for cross-validation of the job analysis survey findings (via longitudinal assessment -- we'll talk about that next!)

Content area vs. item-level relevance ratings

Leveraging Longitudinal Assessment Data

- In 2018, the American Board of Pediatrics (ABP) launched its longitudinal assessment (MOCA-Peds)
 - Short assessment (~15 items) every 3 months
 - Administered via personal device (unproctored)
 - Immediate performance feedback
- Participants are asked to rate the relevance of each item



“How relevant was this question to your practice?”

- Not at all relevant (1)
- Slightly relevant (2)
- Moderately relevant (3)
- Very relevant (4)

Item level confidence and relevance ratings

GENERAL PEDIATRICS WELCOME DR. VALERIE HAIG · ABPID#: 1239506

Question

A 6-week-old exclusively breastfed infant has had bloody stools since that time his mother reported that he was not eating well except for an increase in the frequency of his stools. She was advised to eliminate cow milk products from his diet. She returns for evaluation today because the infant has not breastfed well and his fussiness, blood and mucus daily. A careful history of his diet excludes mammalian milk products in the last 4 weeks.

The infant's examination reveals a normal weight and length, genitalia, and no evidence of an acute illness.

Which of the following is the best management strategy?

- Change the infant's diet to a partially hydrolyzed formula
- Eliminate soy from the mother's diet
- Order IgE testing for food allergies
- Perform a Meckel scan

Submit Answer

By clicking the "Submit Answer" button you are certifying your compliance with the MOCA-Peds Participation Policy.

Question: Confidence

You have selected:

Start a 4- to 8-week trial of acid suppression therapy.

Why am I being asked this?
Your responses to these questions *do not* affect your score. Your responses to these questions *do* affect your score. A summary of your confidence and relevance ratings is available on the My Performance page.

How CONFIDENT are you that you answered this question correctly?

Not at all Slightly

How RELEVANT is this question to your practice?

Not at all Slightly

Submit Responses

THE AMERICAN BOARD of PEDIATRICS

My Portfolio Contact Alerts My Profile

The answer selected is: **INCORRECT**

- Change the Infant's diet to a partially hydrolyzed formula
- Eliminate soy from the mother's diet
- Order IgE testing for food allergies
- Perform a Meckel scan

Exam Domain:
Allergy and Immunology

Learning Objective:
Differentiate between food allergy and food sensitivity.

Rationale:
Food protein-induced allergic proctocolitis (FPIAP) is a non-IgE mediated food allergy that results in inflammation of the distal colon and is most often due to cow milk or soy protein hypersensitivity. This condition typically presents within a few months after birth and can begin as early as the first week after birth. The majority of infants with the condition are exposed to the offending protein via the mother's breast milk while the remainder are formula-fed. Infants with this condition present with blood-streaked stools but are typically otherwise healthy and thriving, as is the case in this vignette. There is no diagnostic test for the condition, but elimination of the responsible allergen typically results in dramatic improvement of symptoms within 72 hours. However, it may take up to a month for complete clearing and resolution of the mucosal lesions.

For this patient, the failure of the symptoms to completely resolve with elimination of cow milk from the mother's diet suggests that he may be sensitive to another protein. Approximately 40% of infants with FPIAP will have sensitivity to both cow and soy protein. Therefore, the next best step is elimination of soy from the mother's diet.

Changing to a partially hydrolyzed cow milk formula would not be appropriate as these formulas have a sufficient amount of intact protein and would continue to cause symptoms. Transition to an extensively hydrolyzed formula would be an appropriate change for a formula-fed infant. Because FPIAP is not an IgE-mediated disorder, IgE testing is not indicated. A Meckel scan is an

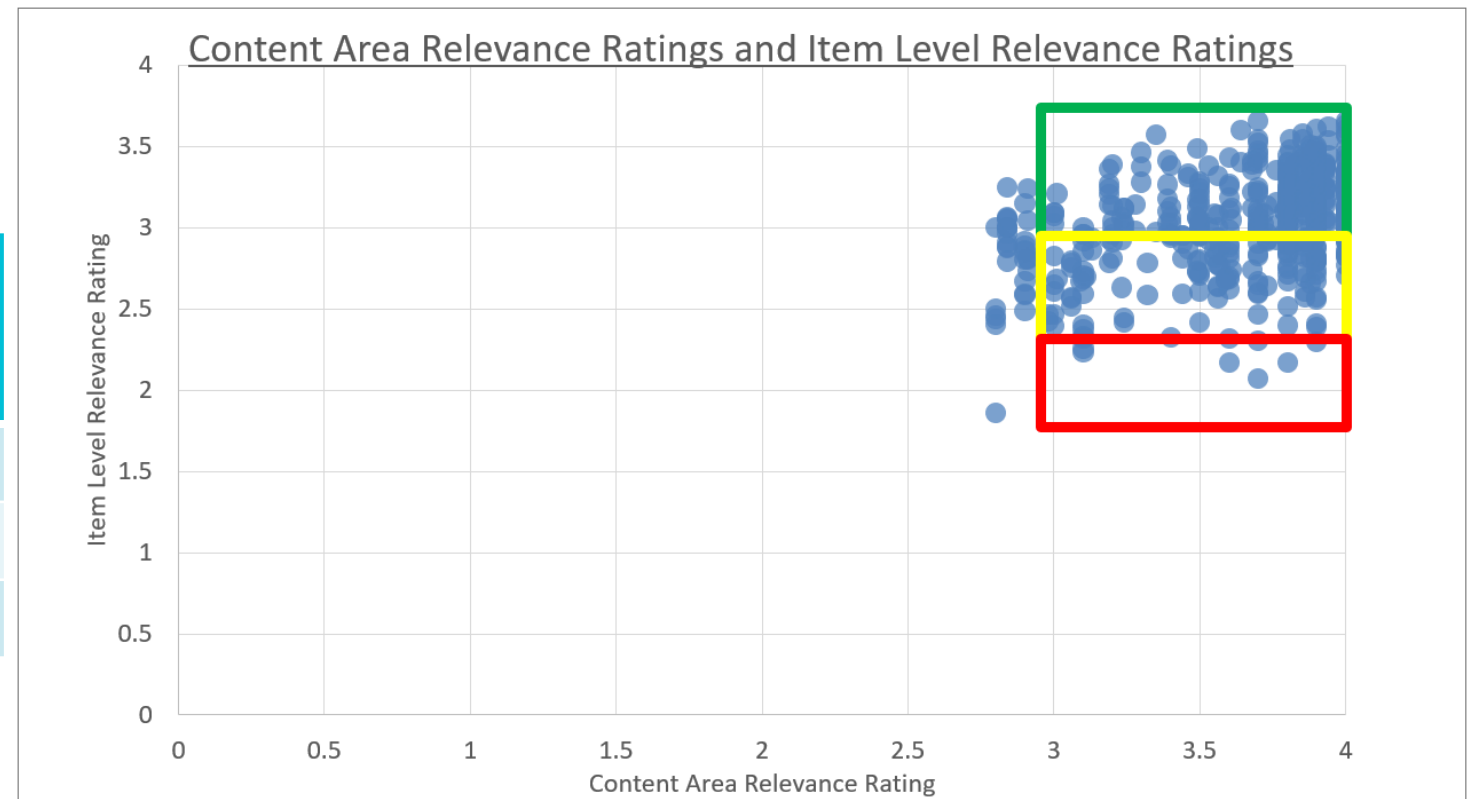
© 2009 -

Question

- Do item-level relevance ratings obtained through longitudinal assessment correlate with the content area relevance ratings obtained through the job analysis survey?

Analysis results

Subspecialty	N Items	Correlation (item vs content area)
Cardiology		0.48
Endocrinology		0.38
Rheumatology		0.44



Future Directions

- Investigate differences between item-level and content category ratings
- Provide relevance data to item writers as a feedback mechanism
- Use item-level relevance ratings as “real time” job analysis data

NCARB case study: Expanding stakeholder input and engagement

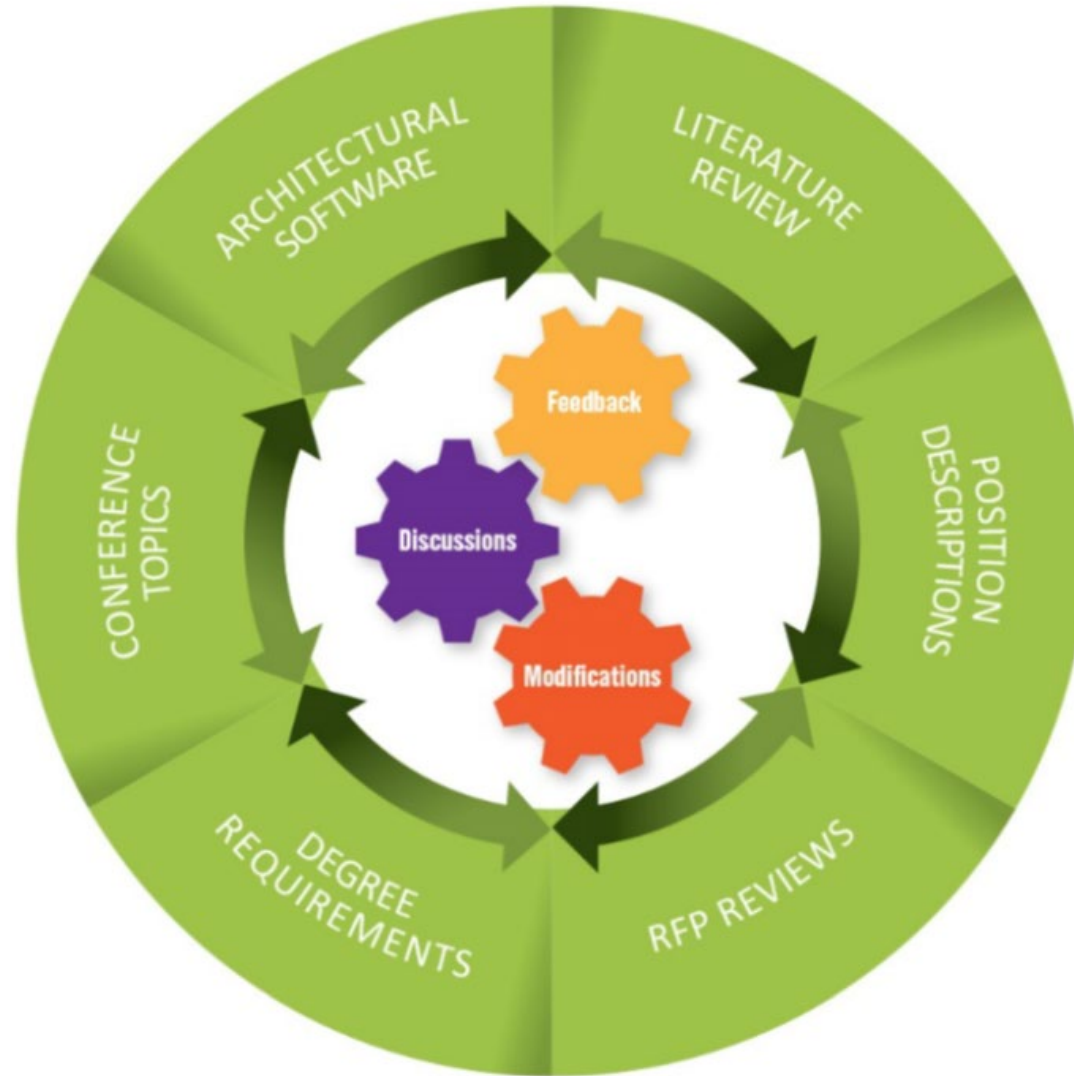
NCARB Analysis of Practice

- Goals:
 - Breadth and depth of information
 - Maximize participation
- Strategy
 - Move beyond Focus group -> Validation Survey
 - Diversified data collection mechanisms -> Diverse stakeholders

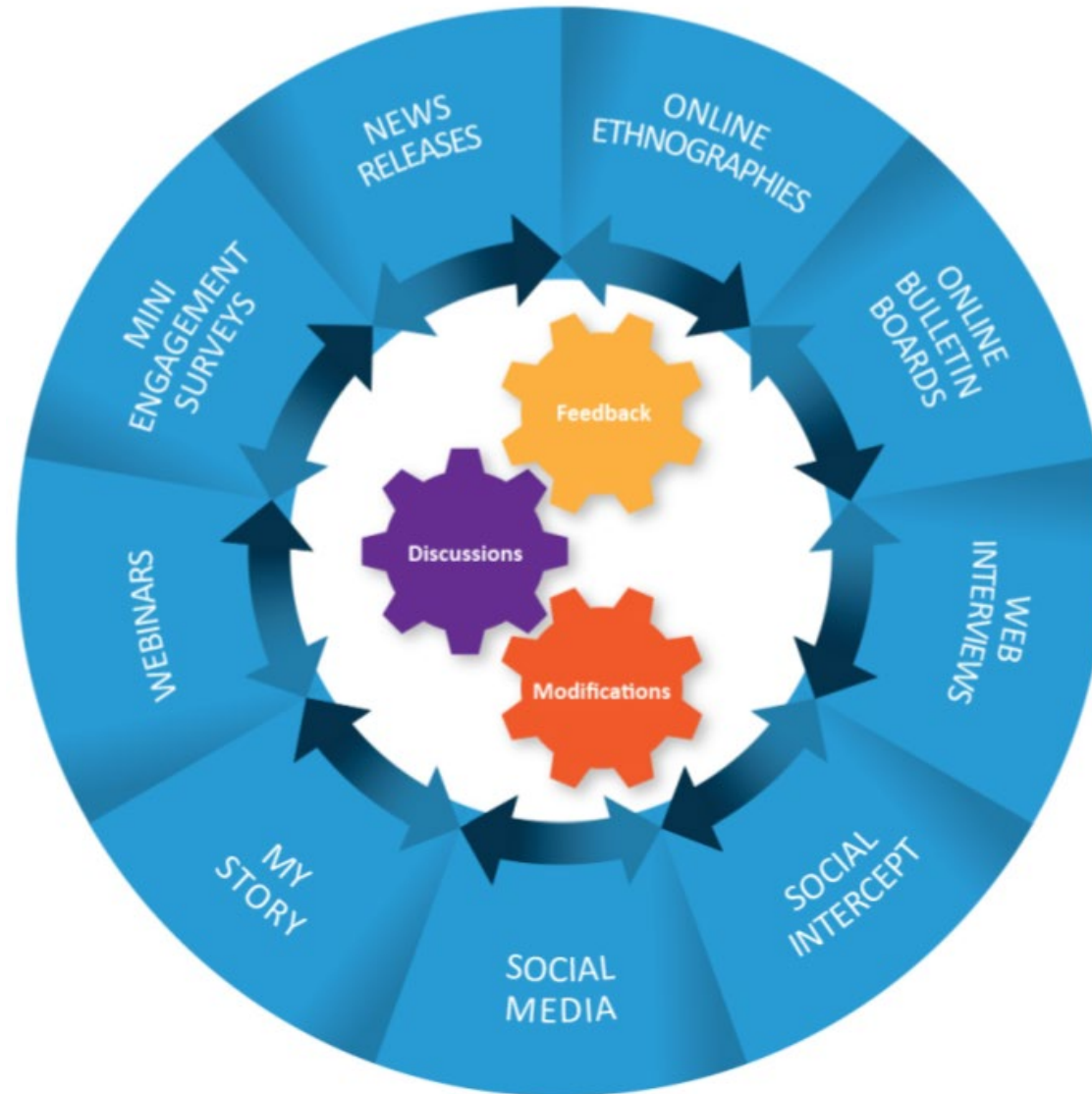
- PHASE I – Qualitative Data Collection
- PHASE II – Mixed Methods for Data Collection and Validation
- PHASE III – Synthesis and Continued Validation



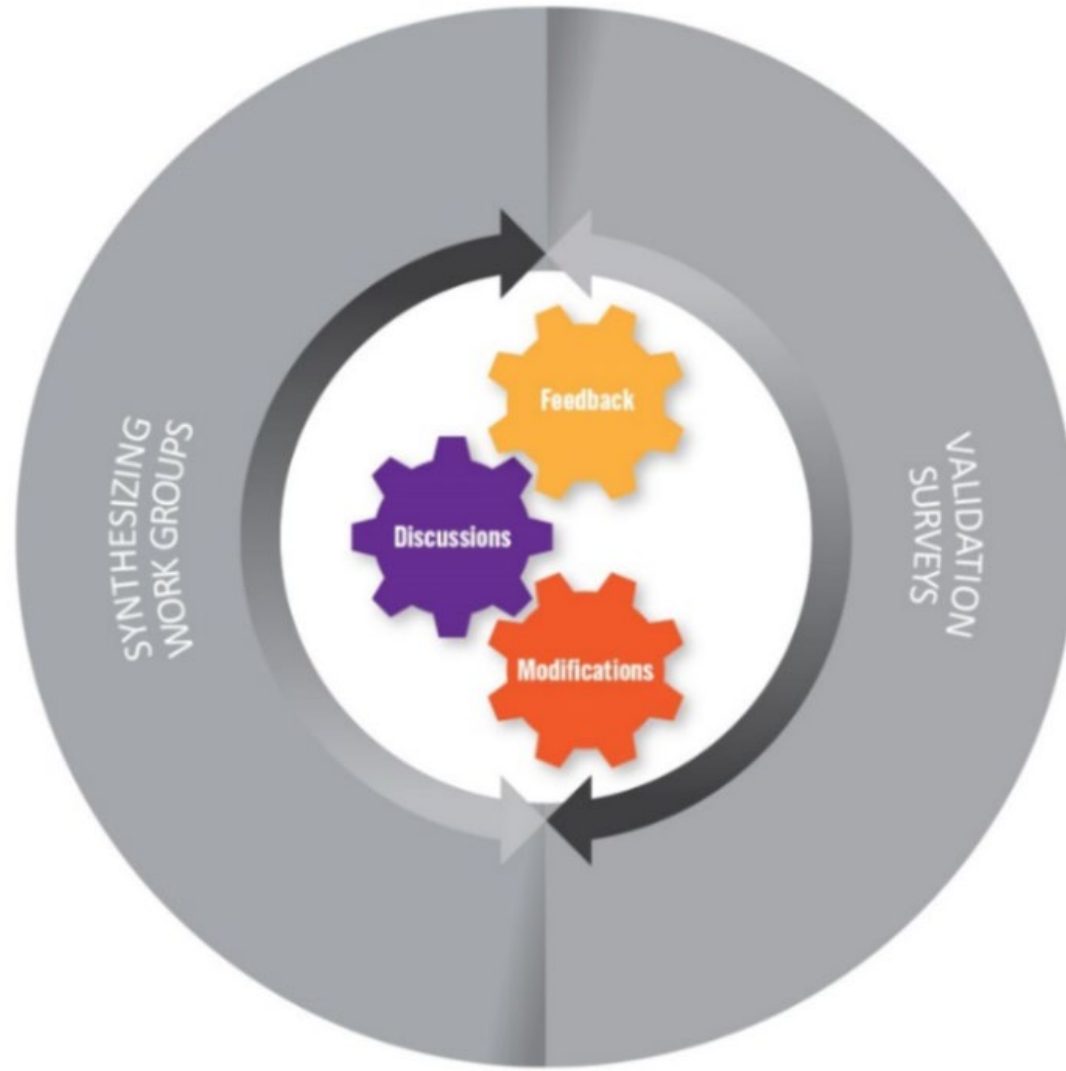
PHASE I












PHASE II



PHASE III



	Method	Symbol	Sample Size
Small Group Activities	Background Research (Phase I)		27
	Ethnographies		28
	My Story		83
	Online Bulletin Board		26
	Social Intercepts		72
	Web Interviews		45
	Webinars		131
Large Group Activities	Mini Engagement Surveys		8,939
	Social Media		1,970
TOTAL			11,321

Phase	Method	Sample Size
I	Background Research	27
II	Small Group Activities	385
	Large Group Activities	10,909
III	Participants ¹	13,446
TOTAL²		24,767

ChatGPT and job analysis possibilities

ChatGPT

- AI as a starting point

Critical and Frequent Tasks

*Pretend you are an expert in the credentialing of pediatricians. Give me a list of the 20 most **critical** tasks that a pediatrician would perform in their practice.*

*Then give me a list of the 20 most **frequent** tasks that pediatricians perform in their practice. Put your response in a two-column table, one column for criticality, one for frequency.*



While I can provide a general idea, keep in mind that I'm not a medical expert. The specific tasks that pediatricians perform can vary widely depending on the healthcare setting, the geographic location, the resources available, and the specific needs of the population they serve.

Most Critical Tasks	Most Frequent Tasks
1. Neonatal resuscitation	1. Routine check-ups
2. Emergency trauma care	2. Immunizations
3. Managing severe infections	3. Developmental screenings
4. Diagnosis of rare diseases	4. Prescribing medication
5. Surgical consultations	5. Treating common illnesses
6. Intensive care unit support	6. Educating parents
7. Managing chronic conditions	7. Referring to specialists
8. Treating congenital anomalies	8. Follow-up appointments
9. Conducting complex diagnostic tests	9. Allergy tests
10. Interdisciplinary case management	10. Ear, nose, throat checks

Identify Domains

Take the 40 tasks you listed above and group them into 4-6 meaningful categories.

Repeat this process 3 times grouping them different ways.

Categorization Method 1: Based on Patient Health Status

Preventive
Routine

Routine

Immunizations

Developmental
screening

Treating
illnesses

Follow-up

Categorization Method 2: Based on Skill Set Required

Clinical

Routine
check-ups

Treating
common
illnesses

Follow-up
appointments

Height
checks

Conducting
physical
exams

Prescribing

Categorization Method 3: Based on Age Group

Infant Care	Child Care	Adolescent Care	General Tasks
Neonatal resuscitation	Treating common illnesses	Managing chronic conditions	Routine check-ups
High-risk newborn care	Allergy tests	Asthma management	Immunizations
Follow-up appointments	Blood tests	Behavioral assessments	Developmental screenings
Conducting physical exams	Treating congenital anomalies	Management of severe mental health issues	Educating parents
Height/weight checks	Fever management	Genetic counseling	Prescribing medication
Immunizations	Vision tests	Counseling on nutrition	Referring to

Blueprint Development

Take the 40 tasks identified above. Create a test blueprint using those tasks. Test should be 150 items.

The number of items related to each task should be proportional to that task's combination of criticality and frequency.



Creating a test blueprint involves identifying the key competencies being tested and allocating test items in a manner that reflects the importance and frequency of each competency in practice. A test blueprint helps to ensure that the assessment tool provides a balanced and representative evaluation of the skills and knowledge being assessed.

Given that there are 150 test items, and considering both the criticality and frequency of the tasks, I'll allocate the number of test items proportionally. Some tasks will be represented more than others, particularly if they are both critical and frequent in practice.

Test Blueprint

Preventive and Routine Care (Total Items: 35)

1. Routine check-ups - 8 items
2. Immunizations - 6 items
3. Developmental screenings - 6 items
4. Treating common illnesses - 4 items
5. Follow-up appointments - 4 items
6. Height/weight checks - 3 items
7. Counseling on nutrition - 2 items
8. Asthma management - 1 item
9. Vision tests - 1 item

Chronic and Long-term Management (Total Items: 30)

1. Managing chronic conditions - 8 items
2. Intensive care unit support - 6 items
3. Medication management in polypharmacy - 5 items
4. Parental education for complex conditions - 4 items
5. Genetic counseling - 3 items
6. Management of severe mental health issues - 2 items
7. Specialized treatment plans for conditions like cystic fibrosis - 2 items

Acute and Critical Care (Total Items: 35)

1. Neonatal resuscitation - 8 items
2. Emergency trauma care - 7 items
3. Managing severe infections - 6 items
4. Treating congenital anomalies - 4 items
5. Treating severe allergic reactions - 3 items
6. Ethical decision-making - 3 items
7. End-of-life care - 2 items
8. High-risk newborn care - 2 items

Diagnostic and Assessment (Total Items: 30)

1. Diagnosis of rare diseases - 6 items
2. Conducting complex diagnostic tests - 5 items
3. Surgical consultations - 4 items
4. Interdisciplinary case management - 4 items
5. Diagnosing pediatric cancers - 3 items
6. Allergy tests - 3 items
7. Blood tests - 2 items
8. Fever management - 2 items
9. Behavioral assessments - 1 item

Communication and Education (Total Items: 10)

1. Educating parents - 4 items
2. Referring to specialists - 3 items
3. Counseling on nutrition - 2 items
4. Parental education for complex conditions - 1 item

Miscellaneous (Total Items: 10)

1. Prescribing medication - 4 items
2. Conducting physical exams - 3 items
3. Ear, nose, throat checks - 2 items
4. Flu tests - 1 item

Convert Tasks to Test Objectives

Take the test blueprint and rewrite the tasks as test objectives that could be assessed using multiple choice items written at a moderate to high cognitive complexity level (i.e., assessing the test taker's ability to analyze, evaluate, apply, or understand).



Certainly! I will reframe the tasks into test objectives suitable for moderate to high cognitive complexity level assessment, using multiple choice items:

Task	Test Objective
Life-saving emergency interventions	Evaluate a candidate's ability to determine the appropriate emergency response to various pediatric health crises.
Diagnosing rare or complex illnesses	Analyze given symptoms and patient histories to correctly diagnose uncommon pediatric conditions.
Neonatal resuscitation	Assess the test taker's understanding of the necessary steps and precautions during neonatal resuscitation.
Routine physical examinations	Analyze and interpret the results of a physical examination to identify any abnormalities in a pediatric patient.
Administering immunizations	Evaluate the correct timing, dosage, and procedure for administering various immunizations to pediatric patients.
Treating common illnesses like	Assess the ability to recommend appropriate treatments based

Test blueprint guidelines:

Level of detail

Background context

- ABP has 16 certification programs (ie, 16 content outlines)
- Historically, content outlines were very detailed (eg, 150 pages)
- In 2017, we started converting our outlines w/ each job analysis:
 - Much less detail (ie, 8-10 pages)
 - Additional classification schemes for items
- In 2020, we developed a set of guidelines for how new content outlines should be structured and organized

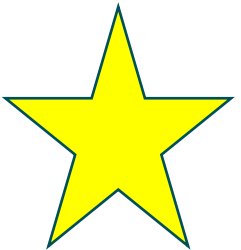
ABP Content Outline Guidelines

Overall Aims:

1. Maximize the content validity of ABP certification exams
2. Maximize the utility and efficiency of test development and psychometric activities

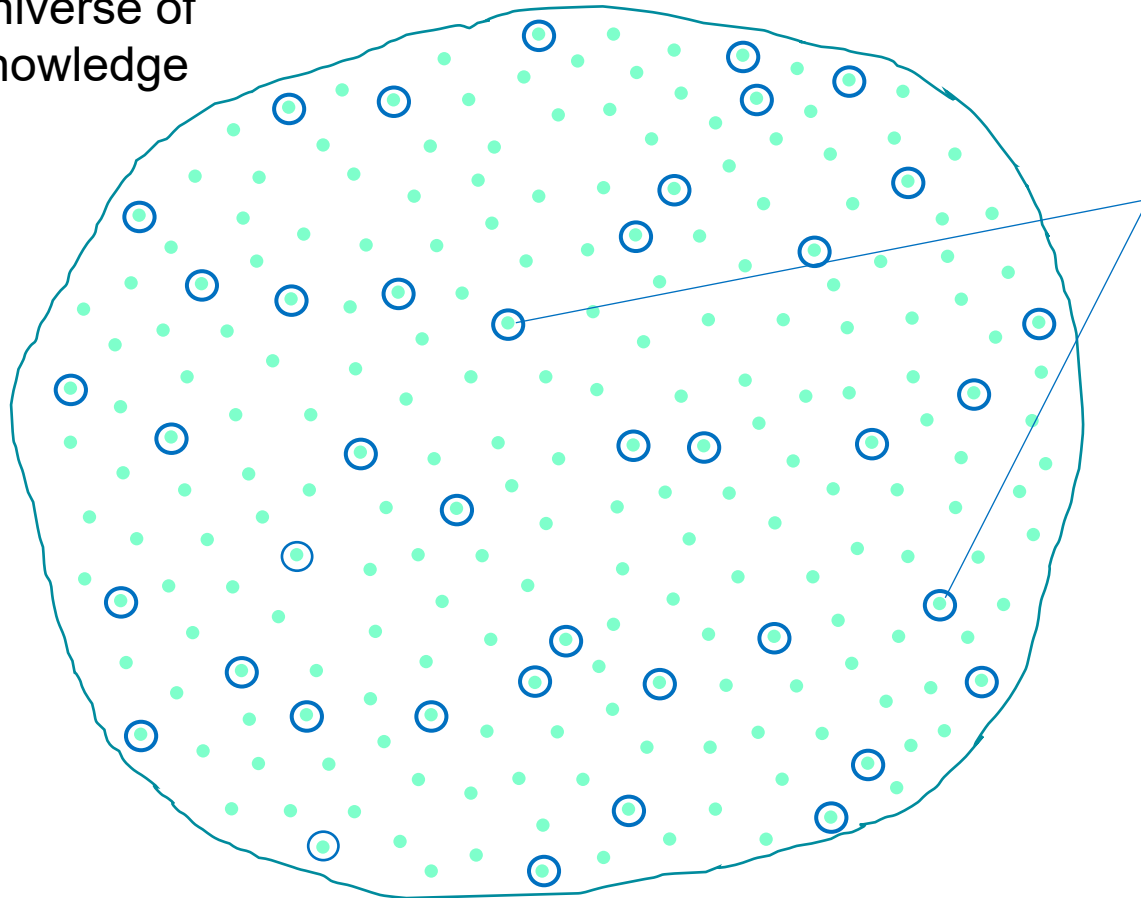
Table of contents

- A. Number of domains and domain weights
- B. Number of subdomains within each domain (ie, level of detail)**
- C. Primary and secondary classification schemes for items
- D. Domain clusters
- E. MOCA-Peds learning objectives



Measurement Model: Sampling from the Universe

Universe of knowledge



Approach:

1. Assess an individual's knowledge on a sample of elements
2. Use knowledge on the sample of elements to make inferences about someone's overall level of knowledge

Objective:

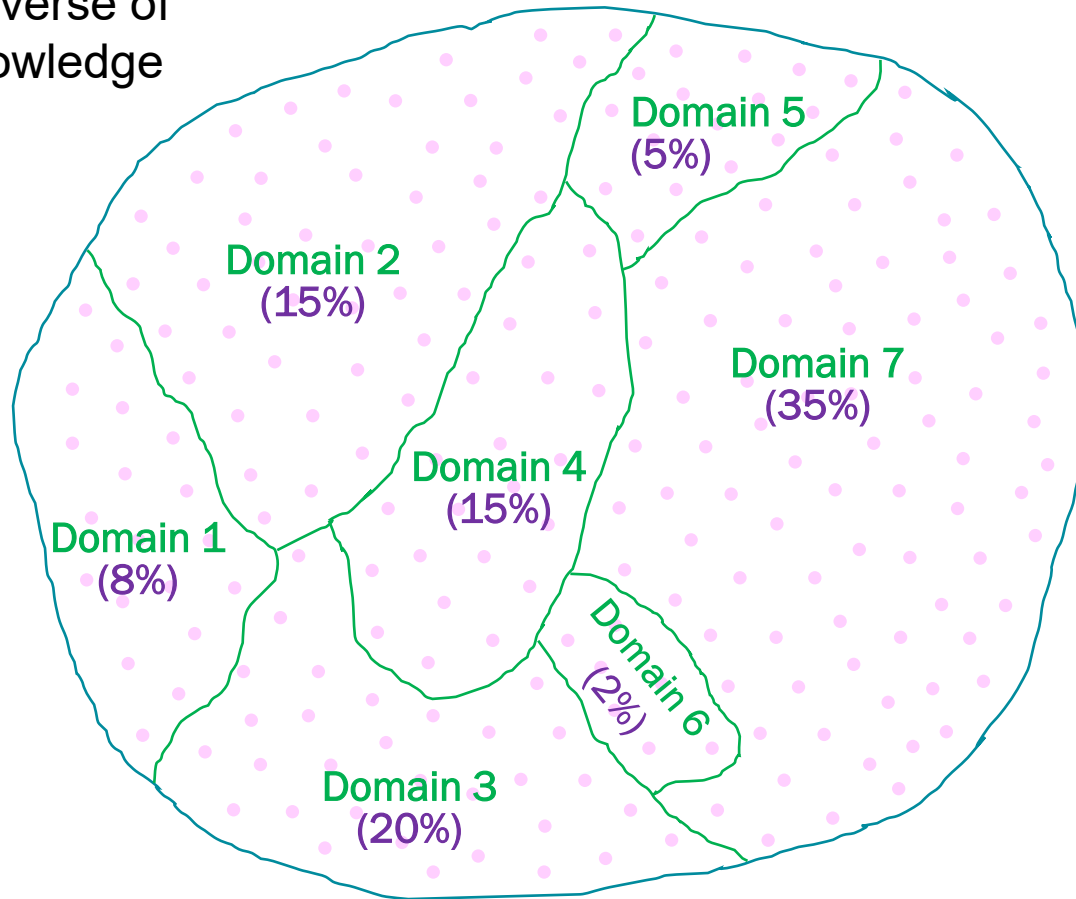
- Valid inferences about an individual's level of knowledge

Sample should be:

- Representative (cover full breadth)
- Focus on most important elements

Sampling philosophy: Categories and weights

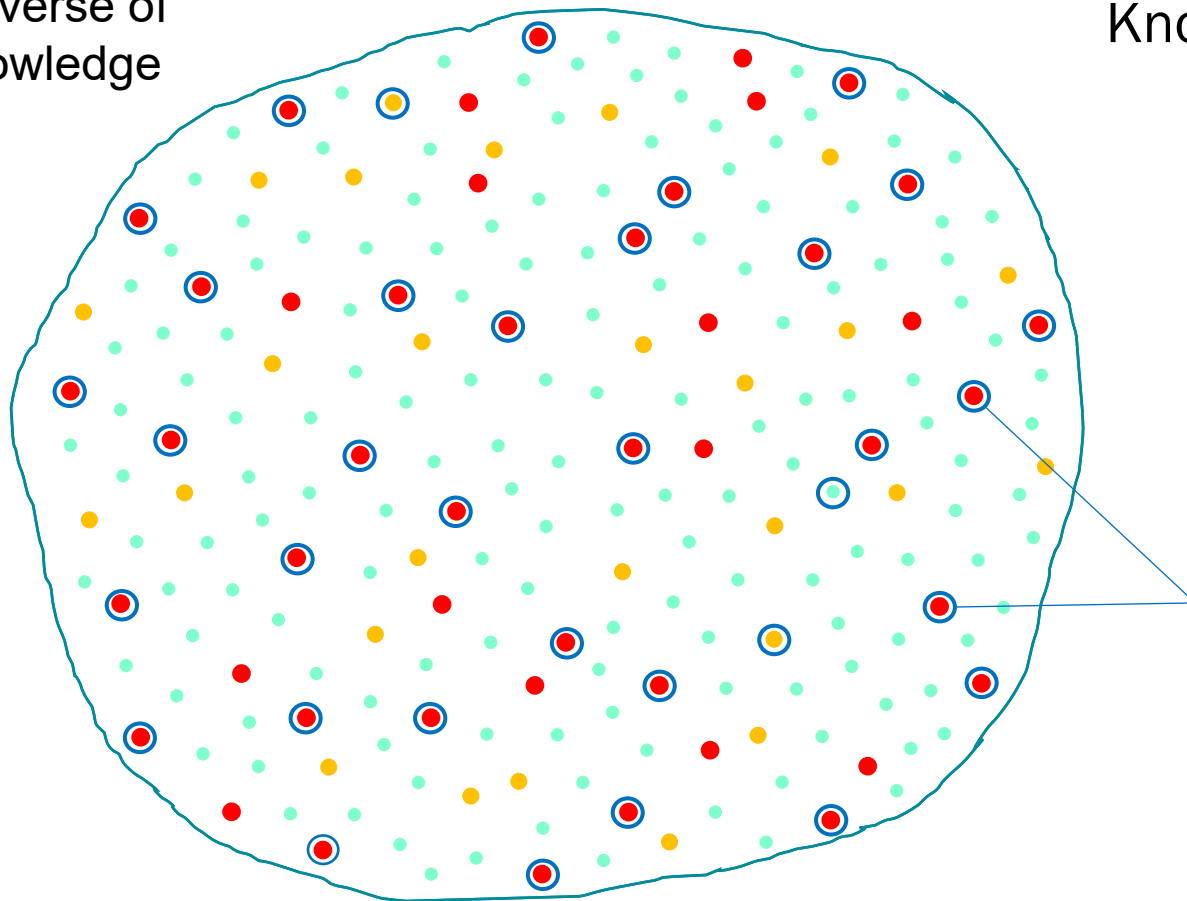
Universe of knowledge



1. Categorize knowledge (into domains, subdomains, etc.) to ensure all important knowledge areas are sampled from
2. Specify weights to ensure the sample reflects the relative importance of each knowledge category

Sampling philosophy: Item level

Universe of knowledge



Knowledge elements NOT equally important

- Highest importance
- Medium importance
- Lowest importance

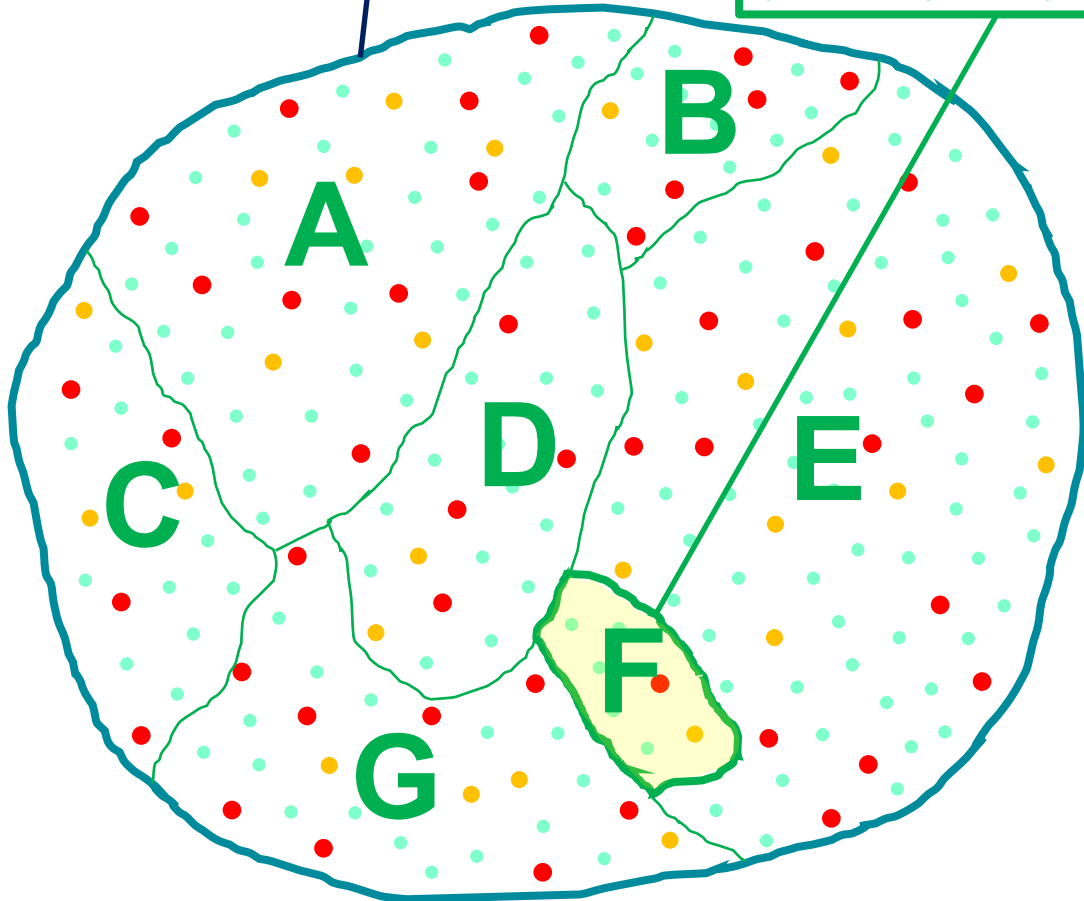
Sampling philosophy:

- To the extent possible, the sample (of items) should assess the most important knowledge elements (within the universe, within a domain, within a subdomain, etc.)

Example

Domain 1:
(10% = 15 items/form)

Subdomain 1.F
(~1 item/form)



1. Domain 1 (15 items)

...

F. Subdomain 1.F (1 item)

1) Sub-subdomain 1.F.1

2) Sub-subdomain 1.F.2

3) Sub-subdomain 1.F.3

4) Sub-subdomain 1.F.4

5) Sub-subdomain 1.F.5

6) Sub-subdomain 1.F.6

Too
much
detail

1. Domain 1 (15 items)

Not
enough
detail

1. Domain 1 (15 items)

...

F. Subdomain 1.F

Just
right!

Level of Detail: Advantages

More detailed outline:

- Provides more specific direction for item writers
- Helps ensures sufficient diversity of items within a domain/subdomain

Less detailed outline:

- More robust to changes in medicine over time (ie, less maintenance)
- Helps to ensure items are assessing the most important elements within a domain/subdomain

Summary

1. The primary purpose of ABP content outlines is to serve as a sampling plan for exam content (ie, not a study guide).
2. Rough rule of thumb – Content outlines should have enough detail so that the lowest level of the outline is expected to have ~1 item per exam form.
 - If the expected # of items per form for a category is > 3 , you may want to add more subcategories
 - If the expected # of items per form for a category is < 0.5 , then you may want to remove lowest level of categories.



Questions?