Continuous Assessment: Psychometric Considerations

Tara M. McNaughton

MEASUREMENT INCORPORATED

EXCELLENCE IN ASSESSMENT

Overview

- Program Design
- Item types
- Feedback
- Formative and summative assessment

MEASUREMENT INCORPORATED

EXCELLENCE IN ASSESSMENT

Terms

- **Module**
  - **Complete assessment from one time point**
- **Cycle**
  - **A series of modules that fully sample the content guidelines/exam blueprint**

Program Design

- Use Current MOC/Recertification exam as preliminary content guideline
- Define a cycle
  - 3 times per year, 3 year cycle
  - 4 times per year, 2 year cycle

Content

- **Review MOC/Recert exam blueprint**
  - **Currency**
  - **Relevance**
  - **Adaptability to continuous assessment**
- **If updating is needed**
  - **Methods/Data to use**
  - **Time to complete vs proposed start of continuous testing**

MEASUREMENT INCORPORATED

Module Construction

- **Complete mini-exams**
- **Content Specific Modules**
  - **1-3 content areas per module**
  - **Vary content-specific modules across diplomates to reflect practice profile**
  - **Consider the balance of content over each year**
    - Face validity
- **Fulfill exam blueprint over the cycle**
  - **Provides validity support**
  - **Necessary for summative assessment**

Program Evaluation

- **Consider on-going program evaluation**
  - Surveys
  - Comments
  - Diplomate performance
- **Diplomates may have suggestions for improvements**
- **Remain flexible to find a good fit**

MEASUREMENT INCORPORATED

Item Types

Multiple Choice Items

- **Most efficient item type**
  - Provide most information about diplomates with the least amount of resources
    - seat time, analysis, etc.
    - Fewer item writing resources
- **Efficiency may no longer be a top concern**
  - Other item types might become more attractive

General Item Types

True/False

- **Other dichotomies**

Multiple-choice (MC)

- **Single answer**
- **Multiple answer (differential diagnosis)**

Complex Multiple choice (K-Type)

- **Combination of choices (A and B)**

Essay

- **Paragraph length (how would you manage this patient)**

Short responses

- **Phrase (what is the most likely diagnosis)**

Matching

- **Series of Premises to match with series of responses**
- **Example: 4 images, 6 diagnoses = 4 possible points for 4 correct matches**

Scenario/Testlet

- **Series of items following a longer expository passage, patient history, labs, imaging, etc.**

Hot-spot

- **Click on the correct location**

Video

- **Candidates click a mouse button when they identify certain key points in the video that they have been instructed to detect.**

Needs someone to grade

MEASUREMENT INCORPORATED

EXCELLENCE IN ASSESSMENT

Key Feature

- **Clinical scenario followed by one or more items**
- **Items may be multiple choice**
  - **One or more answers**
    - differential diagnosis
- **True/False or constructed response**
- **Other item types**

Testlets

- A group of related items
- Develop a number of article-based items as a testlet
- Provide an important article in the field to diplomates and develop a short testlet of items around the article

Other Item Considerations

- **Item attempts**
  - One
  - Two
  - More?
- **Will items be reused?**
  - Only if diplomate got incorrect
  - Every so many years
- **Develop clones or similar item with the same teaching point**
- **How to administer these items for remediation**
  - Differs across diplomates

MEASUREMENT INCORPORATED

Item Development

- Established exam programs have large item banks with items of known quality
  - Items may be edited and improved over multiple administrations
- Continuous testing requires more item development
  - Most if not all items may be new
  - Quality of modules harder to gauge prior to administration

Increase focus on item writers

- **Training**
- **Mentoring**
- **Feedback**
  - **Other subject matter experts**
  - **Expert item writers**
  - **Item statistics**
  - **Diplomate comments and performance**

Scoring and Feedback

- **Item-level Scoring and Feedback**
  - **How much and how detailed**
  - **Feedback for each response item**
    - Only show for selected response
    - All options
  - **When to present item level feedback**
    - Stand alone items
    - Testlet or key feature item groups
- **Performance feedback per module**

Formative vs Summative Assessment

Decide how the continuous assessment will be used for decisions about certification

- **Formative – low stakes, similar to CME**
- **Summative – high stakes, standard/outcome**

Formative vs Summative Assessment

- Can combine goals and use one assessment for both, but formative assessment ***must*** have an effective feedback component
- Item writers are also tasked with developing more effective feedback

Setting a Standard

- To make a summative decision on certification status, a standard must be established
- A variety of methods exist to set standards, but for continuous assessment systems, another layer must be added
- McNaughton and Reyes have proposed a 2-stage standard setting process for continuous assessments

Angoff Example

- Rate the items in each module using the modified Angoff method
- Diplomate scores are either
  - Proficient (meets the standard or higher)
  - Borderline (within 1 to 2 SEMs below)
  - Insufficient (more then 1 to 2 SEMs below)

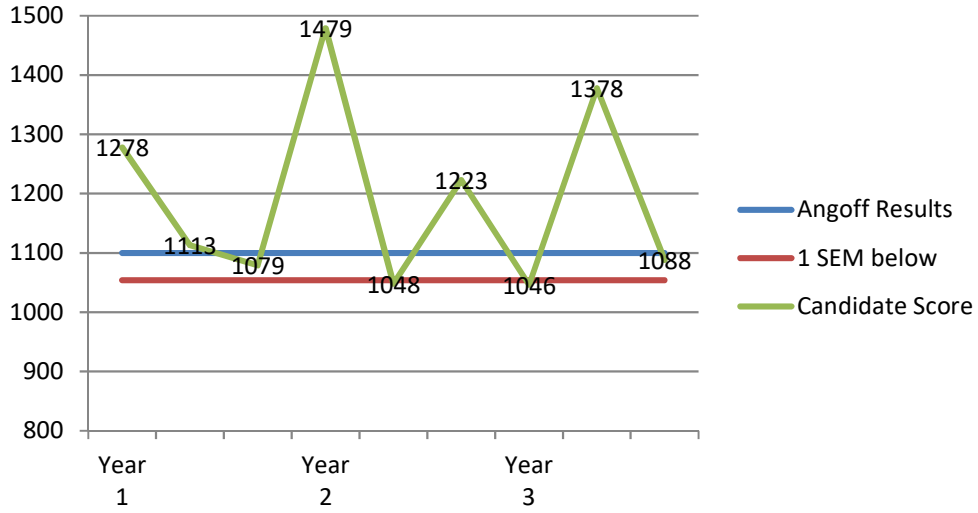
Angoff Example

- Rate items as modules are developed
  - Standards are captured in time
- Continuous testing → continuous feedback to standard setting panel
  - practice makes perfect

MEASUREMENT INCORPORATED

EXCELLENCE IN ASSESSMENT

Summative Judgement



Diplomate Results

5 times - proficient

2 times – borderline

2 times – insufficient

Is this acceptable?

Set Angoff ratings to a pass point of 1100 scaled score points

Second step for the standard:

- How many scores below standard or what proportion of borderline/insufficient
  - Acceptable (retains certification)
- What happens for non-acceptable performance?
  - A probationary period?
- How to deal with missing data? (Missed data points/noncompliance)

MEASUREMENT INCORPORATED

EXCELLENCE IN ASSESSMENT

If Standard Not Met

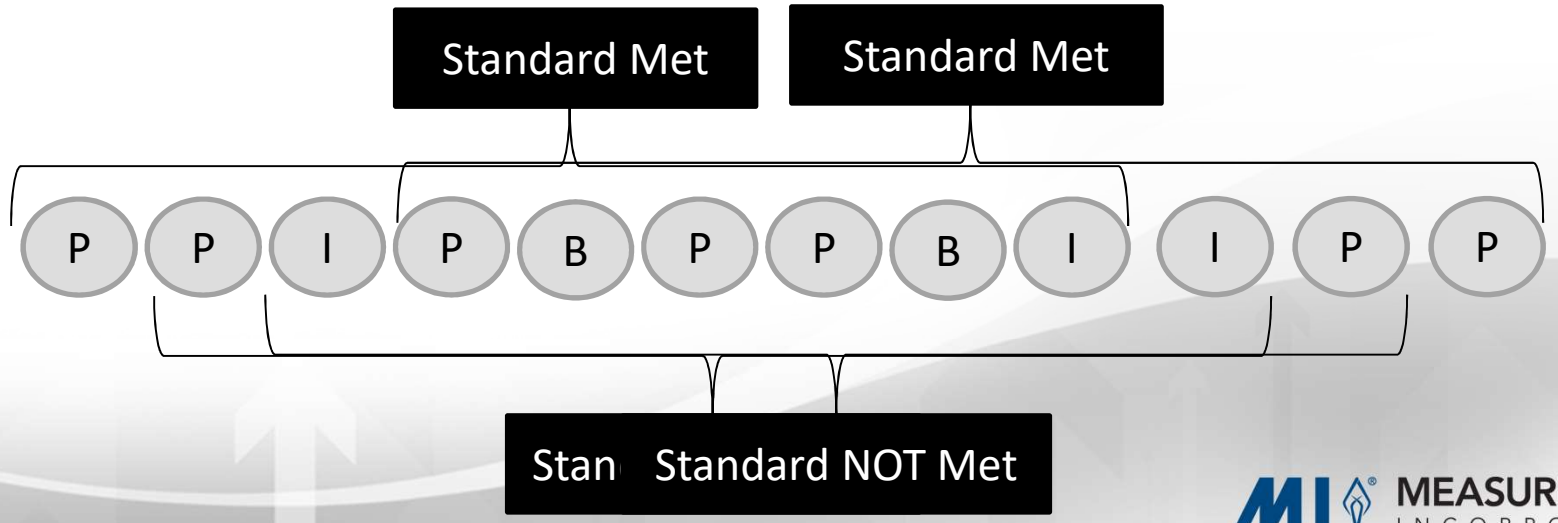
- **Remediate specifically to content not mastered**
  - **Advantage of content-specific modules**
  - **Readminister alternate modules on specific content**
  - **Outside regular module administration**
- **Use the identified cycle**

Rolling Summative Feedback

- **Once the first full assessment cycle is completed, Stage 2 scoring may be applied on an ongoing basis**

Stage 2 standard

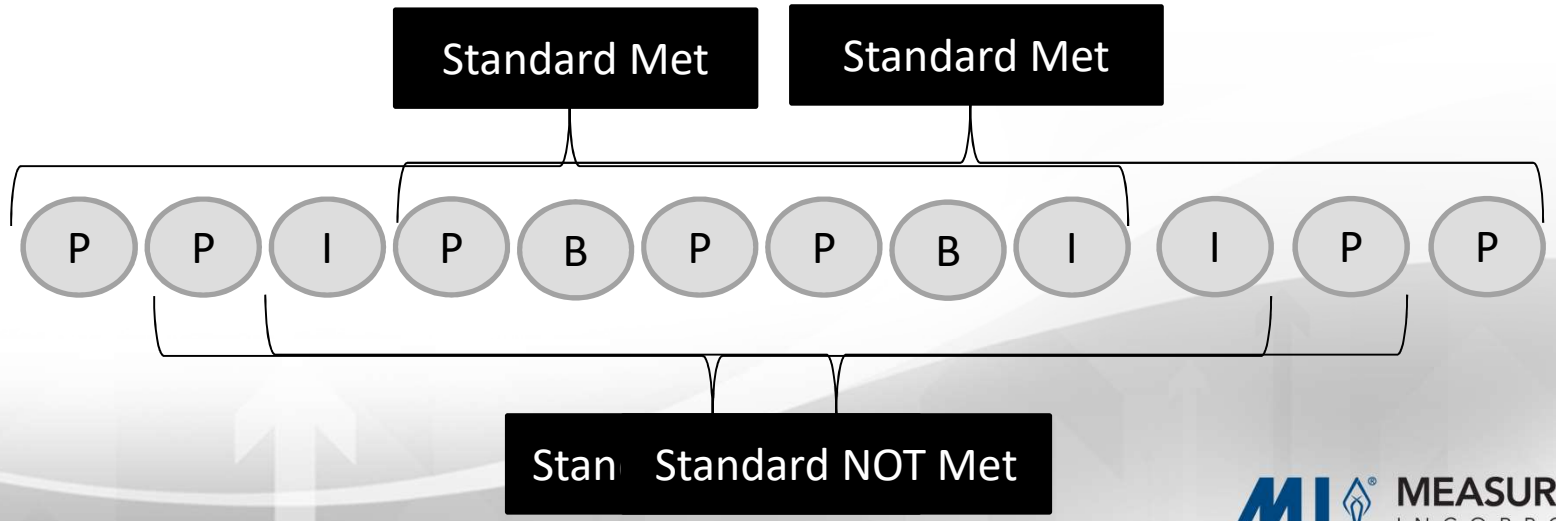
- **Minimum of 5 proficient scores (P)**
- **Maximum of 2 Insufficient scores (I)**



MEASUREMENT INCORPORATED

Summative Feedback

- **Provides a great deal of feedback to diplomates on their performance relative to expectations of competency**



MEASUREMENT INCORPORATED

EXCELLENCE IN ASSESSMENT

Comments or Questions?

Contact Information:

tmcnaughton@measinc.com
(312) 263-9411 x308