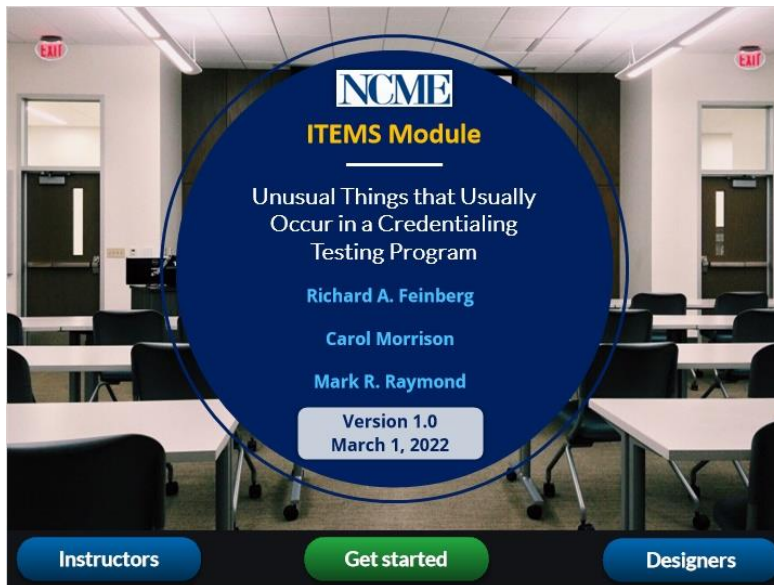


# DM 28: Unusual Things


## 1. Introduction

### 1.1 Module Cover



### 1.2 Learning Objectives

## Learning Objectives



Welcome %LearnerName%!

1. Distinguish different components of the assessment lifecycle
2. Explain different units, roles, and handoffs within a testing organization
3. Recognize potential risks associated with challenges that commonly arise for credentialing testing programs
4. Identify when testing challenges are likely to create validity concerns

### 1.3 Resources

#### Resources

References for more information on this topic :

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Clauser, B.E., Margolis, M.J., & Case, S.M. (2006). Testing for licensure and certification in the professions. In R. L. Brennan (Ed.). *Educational measurement*, 4th edition (pp. 701-731). American Council on Education (ACE)/Praeger.

Lane, S., Raymond, M. R., Haladyna, T. M., & Downing, S. M. (2016). Test development process. In S. Lane, M. R. Raymond, & T. M. Haladyna (Ed.). *Handbook of test development*, 2nd edition (pp. 1-16). Taylor & Francis.

- Also see chapters: 2, 20, 22, 27, 28, 29, and 31.

### 1.4 Main Menu

#### Main Menu

Theory

01 Assessment Lifecycle  
[15 Minutes]

02 Roles within a Testing Organization  
[10 Minutes]

03 Qualifying Risk  
[15 Minutes]

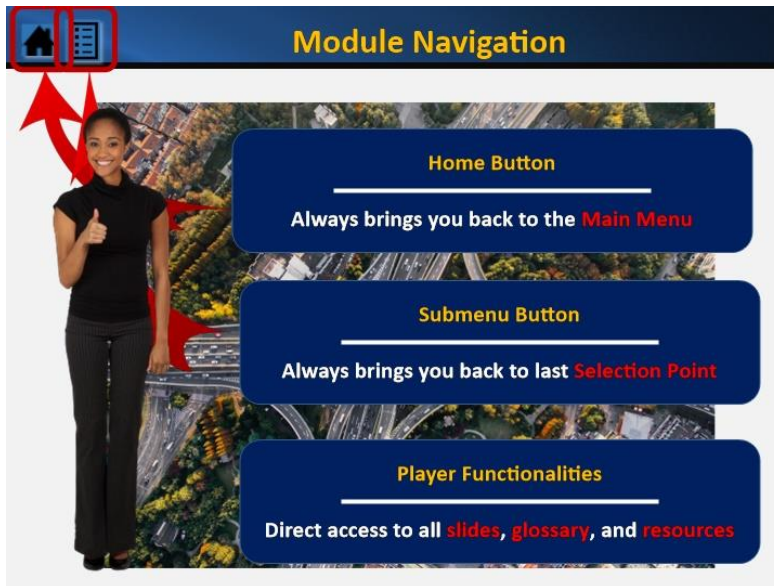
04 Considerations and Action  
[20 Minutes]

05 Interactive Scenarios  
[20 Minutes]

06 Quizzes  
[10 Minutes]

Practice

## Navigation (Slide Layer)






## 1.5 Support



## 1.6 Instructors

Meet the instructors:

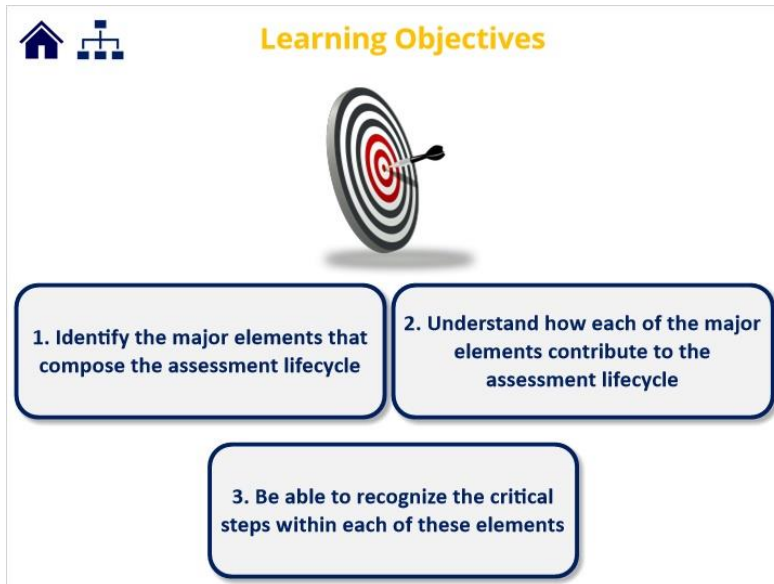
		
<b>Richard A. Feinberg</b> National Board of Medical Examiners	<b>Carol Morrison</b> National Board of Medical Examiners	<b>Mark R. Raymond</b> National Conference of Bar Examiners

## 2. Assessment Cycle

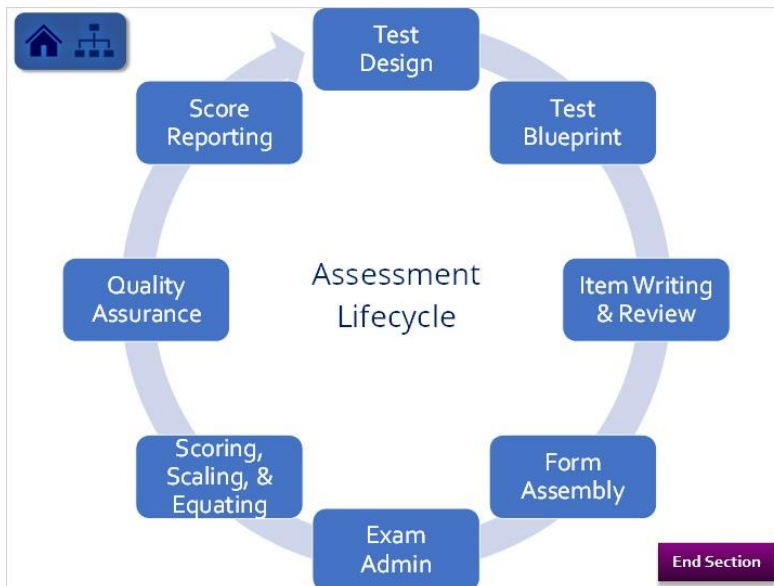
### 2.1 Assessment Cycle Beginning



## 2.2 Learning Objectives: Assessment Lifecycle




## 2.3 Assessment Cycle




## 2.4 Bookmark: Test Design



## 2.5 Test Design



**Test Design**

- A document summarizing the entire assessment lifecycle for an exam problem



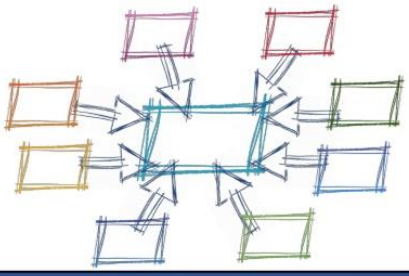


## 2.6 Test Design






### Test Design

- Describes all the key aspects of an exam program such as the purpose, target population, eligibility requirements, mode of administration, and desired inference



## 2.7 Bookend: Test Design





This is the end of this part.



[Back to Assessment Lifecycle](#)

[Next Step in Cycle](#)

## 2.8 Bookmark: Test Blueprint



## 2.9 Test Blueprint





Test Blueprint

- The test blueprint, or test specifications, provides the specific details on the exam material

Content	Psychometric Properties
Format	Form Assembly




## 2.10 Test Blueprint






Test Blueprint

- Informed by a job/practice analysis
- For more information on job/practice analysis see ITEMS Module 23: Practice Analysis Questionnaires: Design and Administration



## 2.11 Bookend: Test Blueprint





This is the end of this part.

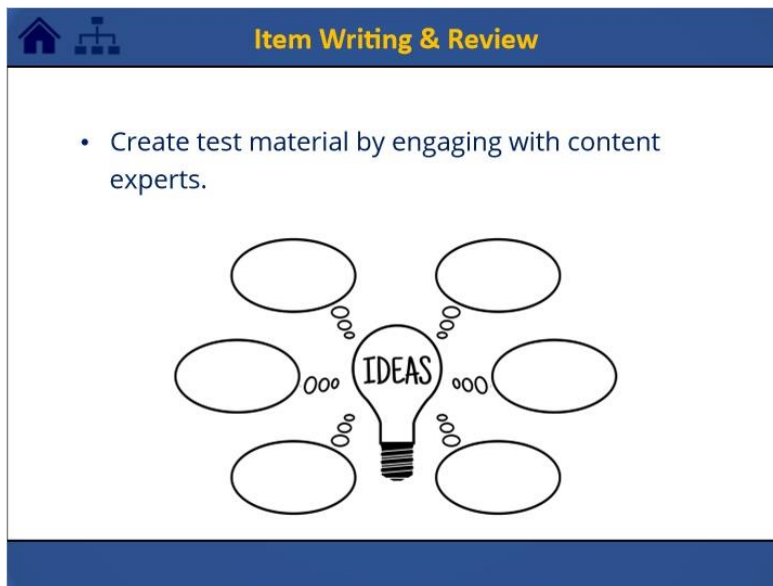
[Back to Assessment Lifecycle](#)

[Next Step in Cycle](#)

## 2.12 Bookmark: Item Writing & Review



## 2.13 Item Writing & Review



## 2.14 Item Writing & Review






### Item Writing & Review

- Reviewing item drafts or older items for validity/ improvement, which may involve using item level performance data



## 2.15 Bookend: Item Writing & Review





This is the end of this part.



[Back to Assessment Lifecycle](#)

[Next Step in Cycle](#)

## 2.16 Bookmark: Form Assembly

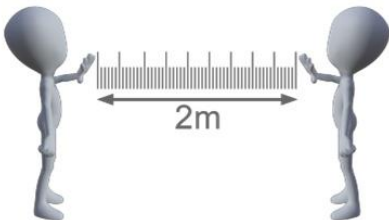


## 2.17 Form Assembly





### Form Assembly

- Building forms according to specifications (e.g., statistical and content) and constraints (item overlap between forms, number of items, item types)

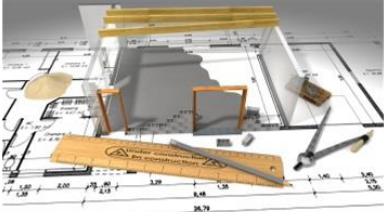
An illustration showing two stylized, grey, humanoid figures standing on a white surface. They are holding a long, horizontal ruler between them. The ruler has markings and the text "2m" is written below it, indicating its length. The entire scene is set against a white background within a blue-bordered frame.

## 2.18 Form Assembly






### Form Assembly

- Form construction is also dependent on delivery mechanisms:
  - Fixed forms
  - Linear-on-the-fly-test (LOFT)
  - Computer adaptive test (CAT)



## 2.19 Bookend: Form Assembly



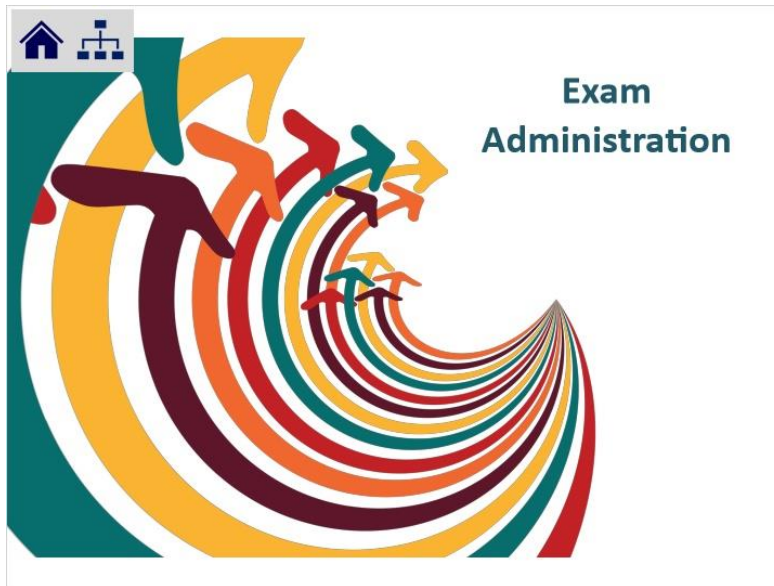


This is the end of this part.



[Back to Assessment Lifecycle](#)

[Next Step in Cycle](#)


## 2.20 Bookmark: Exam Administration



## 2.21 Exam Administration



**Exam Administration**

- Common modalities for exam administration :
  - Computer-based Testing
  - Web-based Testing
  - Paper-based Testing






## 2.22 Exam Administration






### Exam Administration

- Part of exam administration is also to consider how to mitigate threats to score validity:
  - Standardization
  - Timing
  - Accommodations: Extra time, extra breaks, reader, etc.
  - Disruptions



## 2.23 Bookend: Exam Administration

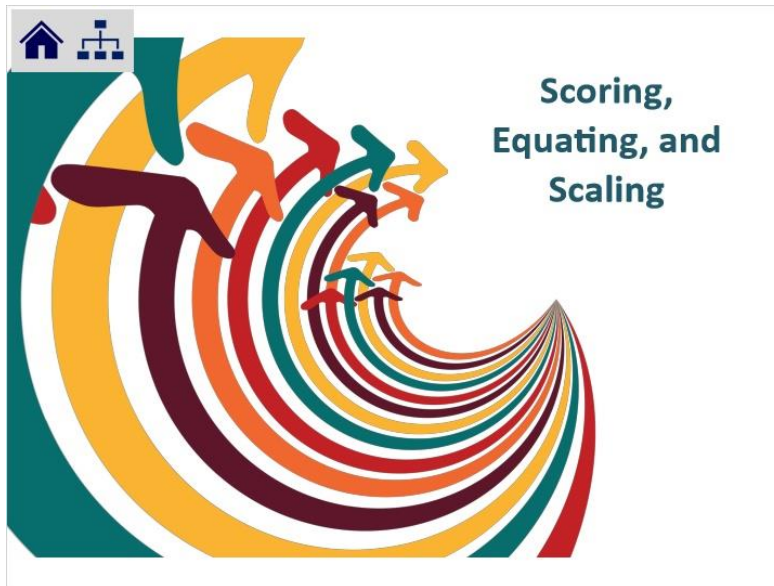





This is the end of this part.

[Back to Assessment Lifecycle](#)[Next Step in Cycle](#)

## 2.24 Bookmark: Scoring, Equating, and Scaling



## 2.25 Scoring, Equating, and Scaling

 **Scoring, Equating, and Scaling**



- After the examination, the next step is to collect the item response data and compute raw and percent correct scores.

**30%**

**50%**


**100%**

## 2.26 Scoring, Equating, and Scaling





### Scoring, Equating, and Scaling

- For exam programs that require comparability across forms either within the same administration or longitudinally, equating is performed to adjust for minor statistical differences between forms
- The right equating method depends on many factors:
  - Sample size of examinees and items
  - Heterogeneity
  - Score inference(s)



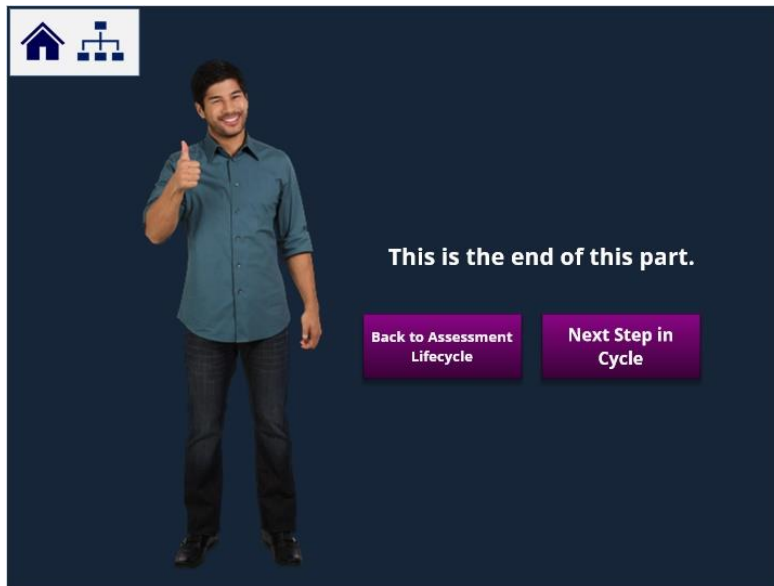
## 2.27 Scoring, Equating, and Scaling



### Scoring, Equating, and Scaling

- Scaling refers to the process of transforming scores to facilitate interpretation for the user
- Scaling can be useful for:
  - Anchoring to a meaningful comparison group
  - Maintaining a fixed value for the passing score over time
  - Comparing relative performance across different tests



## ***2.28 Bookend: Scoring, Equating, and Scaling***



## ***2.29 Bookmark: Quality Assurance***




## 2.30 Quality Assurance





### Quality Assurance

- Quality assurance refers to all the checks following the exam administration
- Validating processes related to:
  - Receiving expected item response data
  - Application of the correct scoring keys
  - All scoring, equating, and scaling procedures
  - Flagging performance for unusual items or examinees




## 2.31 Quality Assurance

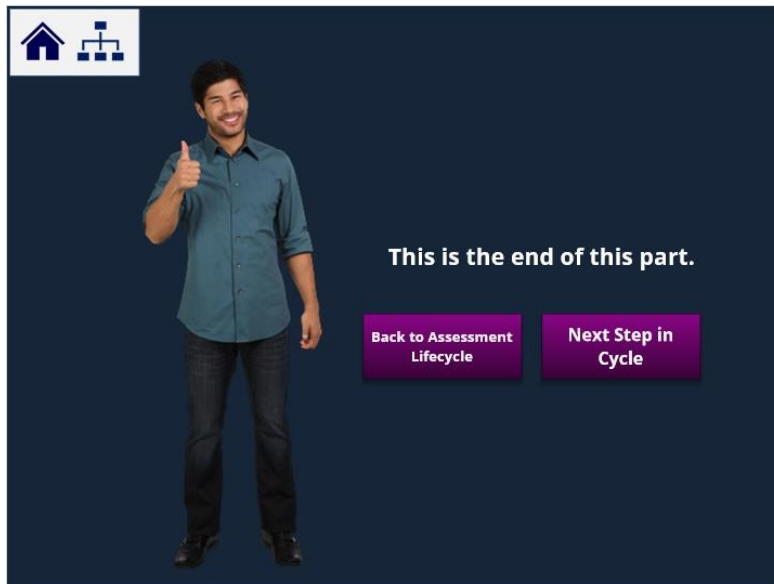


### Quality Assurance

- Quality assurance also includes investigating test day incidents that may be unnoticeable in the item response data
- This can include a proctors report of suspicious behavior or anonymous tips that secure test material is available online



## 2.32 Bookend: Quality Assurance



## 2.33 Bookmark: Score Reporting





## 2.34 Score Reporting



Score Reporting

- Summarizing and presenting results to various stakeholders of an assessment (examinees, schools/programs, governing agencies, licensing boards)



## 2.35 Score Reporting

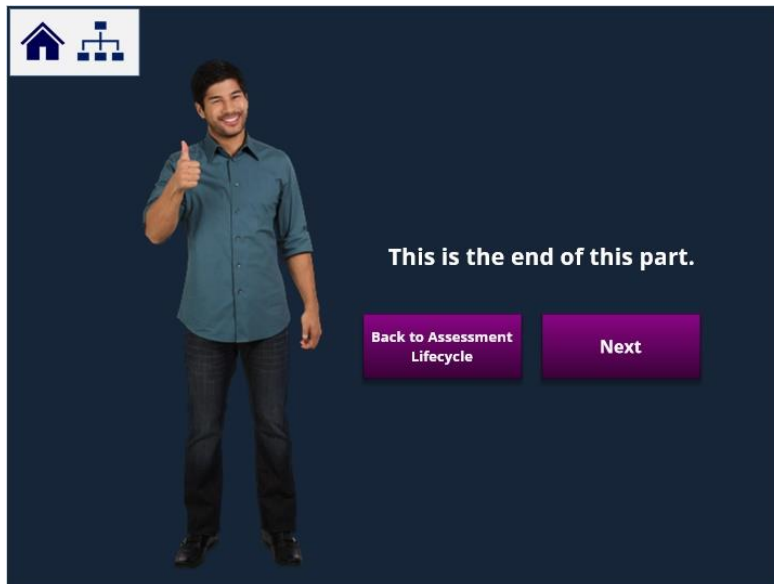


Score Reporting

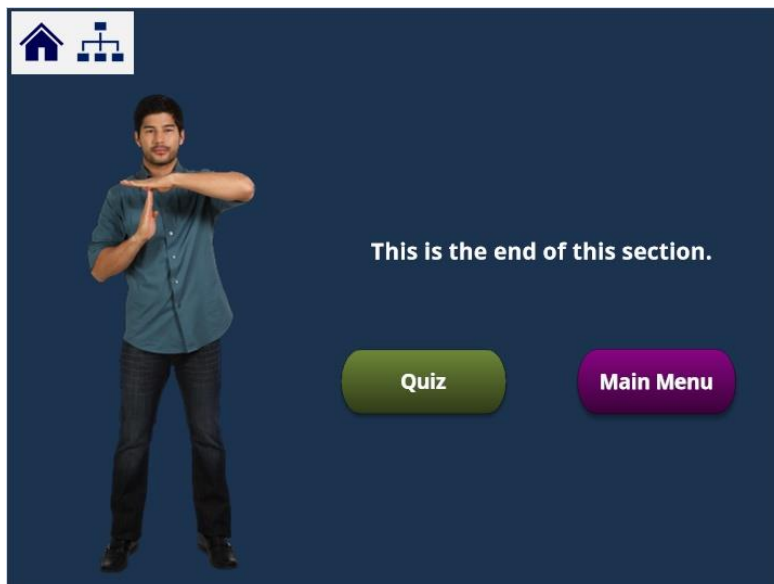
- Score reporting is mainly used as a tool to communicate results:
  - Comparisons to relevant group(s)
  - Performance in a specific content area
  - Feedback for remediation
- How the score information is presented can help encourage examinees and convey value



### ***2.36 Bookend: Score Reporting***



### ***2.37 Bookend: Assessment Cycle***

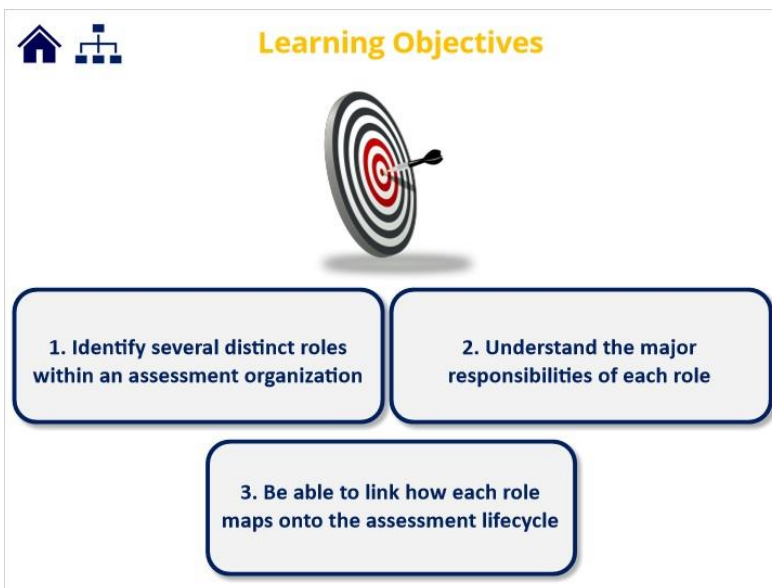


### 3. Roles

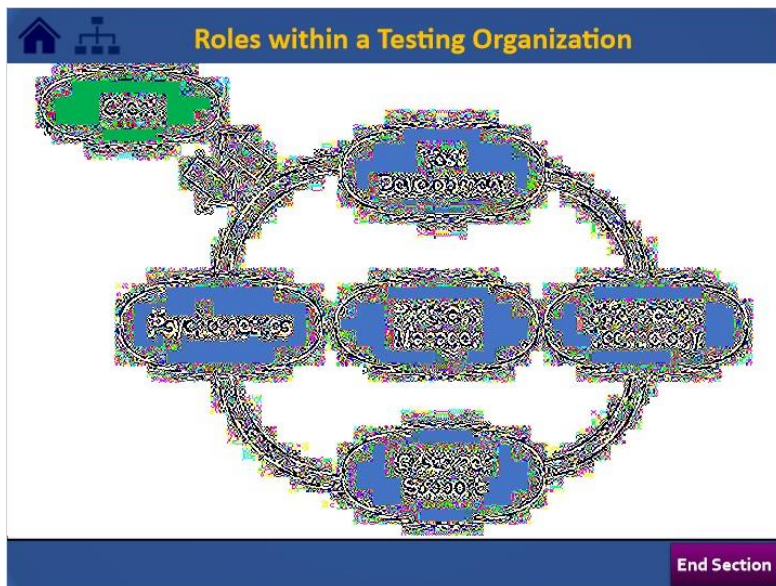
#### 3.1 Roles Beginning



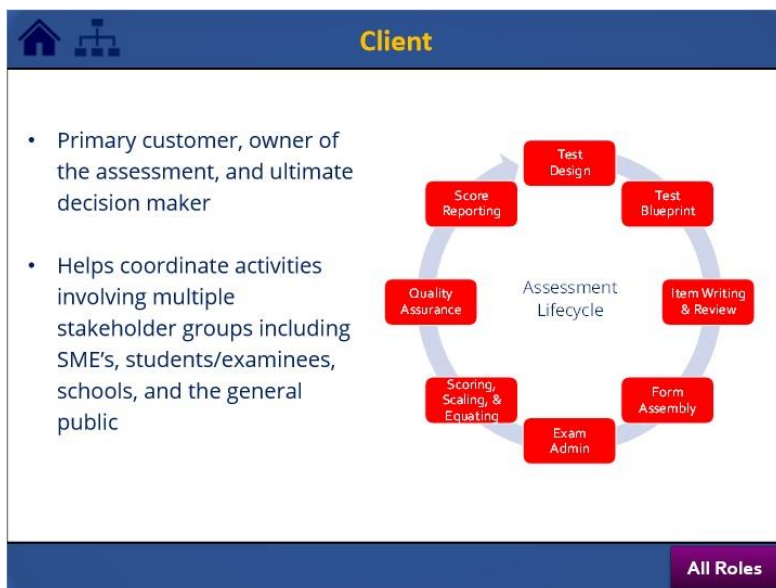
#### 3.2 Learning Objectives: Roles



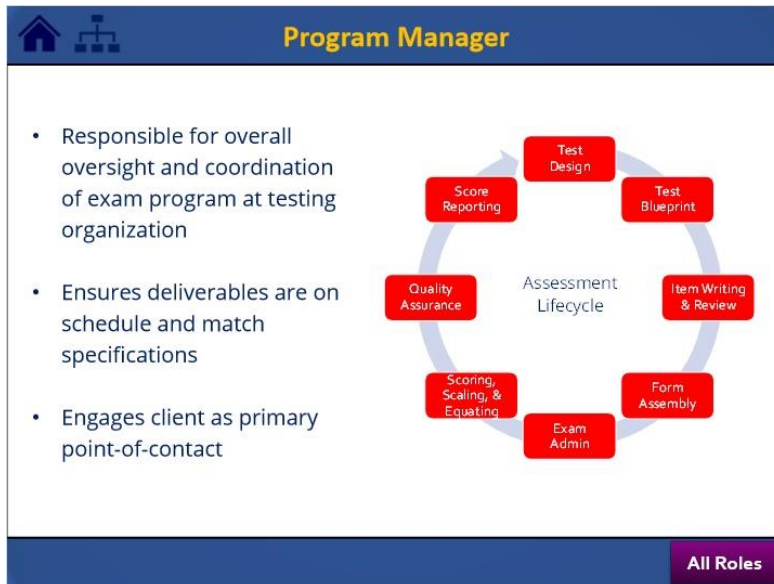
### 3.3 Roles Diagram



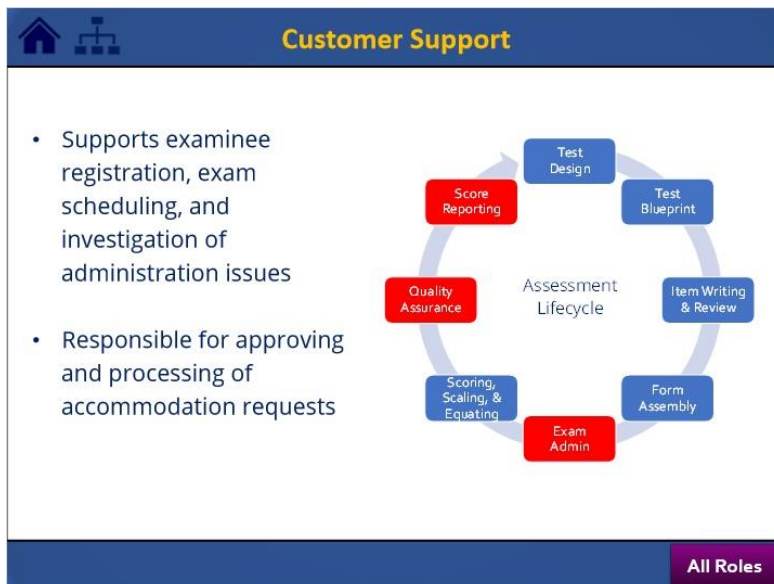
### 3.4 Client



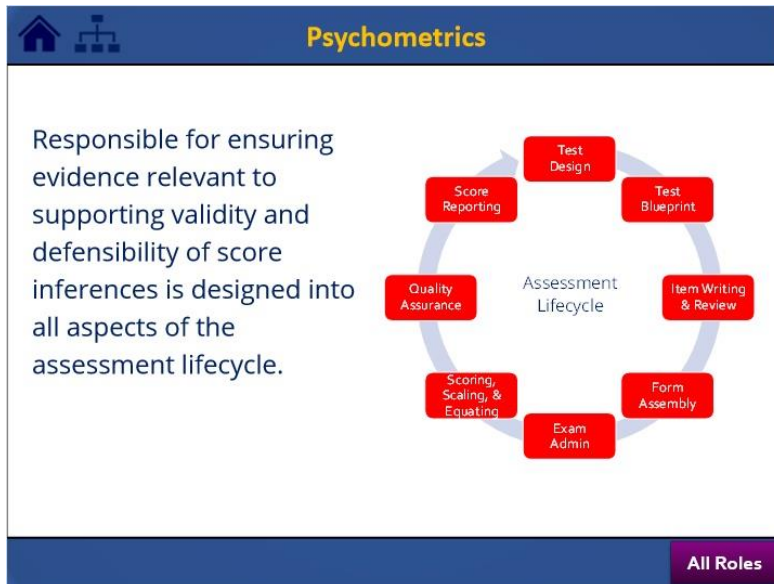
### 3.5 Program Manager



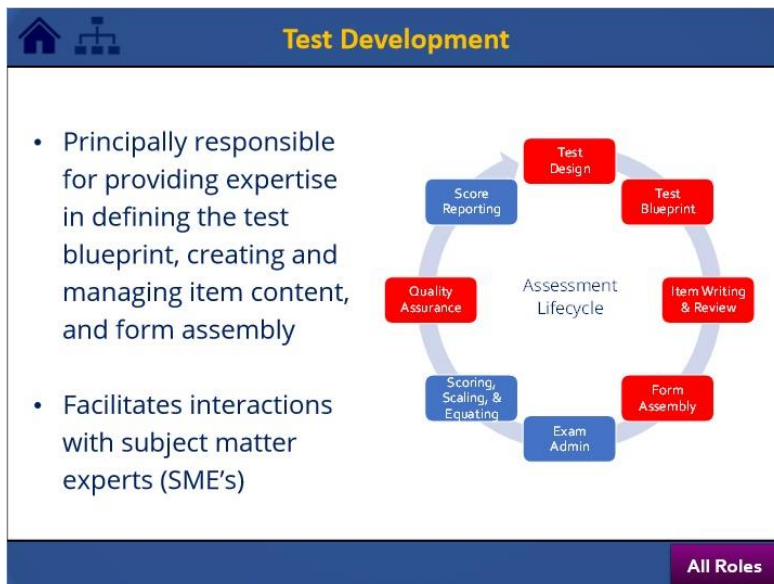
### 3.6 Customer Support



### 3.7 Psychometrics





### 3.8 Test Development






### 3.9 IT



## Information Technology (IT)

Responsible for applying technology to business processes specifically for



- item authoring and banking
- exam delivery
- secure capture of examinee response records
- scoring
- creation of score report materials.




Assessment Lifecycle

All Roles

### 3.10 Bookend: Role





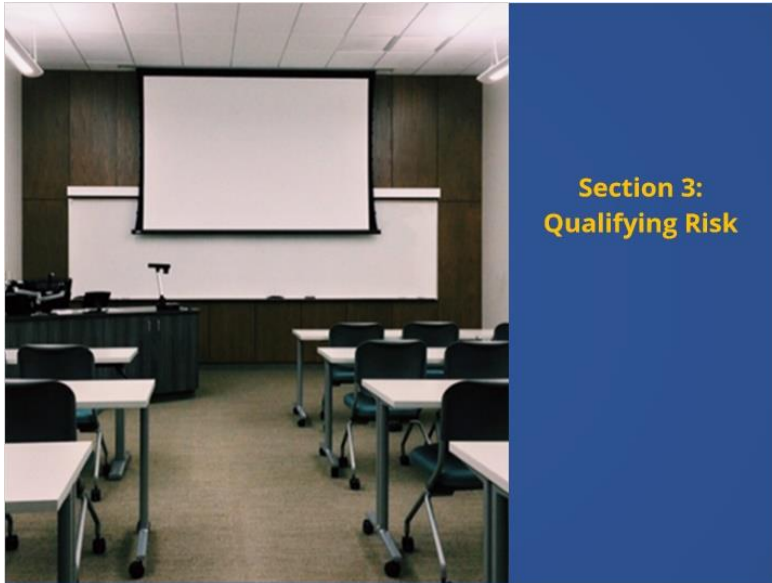
This is the end of this section.

Quiz

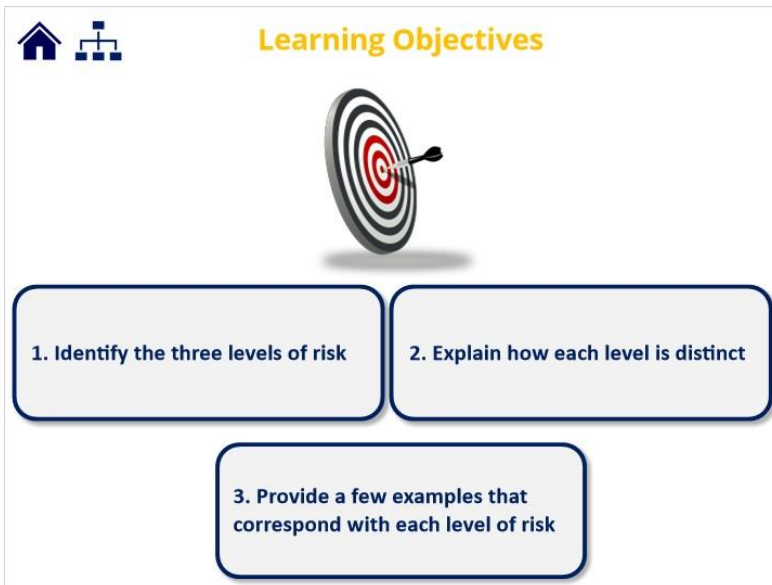
Main Menu

## 4. Qualifying Risk



### 4.1 Qualifying Risk Beginning



### 4.2 Learning Objectives: Qualifying Risk




### 4.3 Categories of Risk



#### Categories of Risk

Can be defined in terms of threat to score validity and are likely subjective relative to the specifications/stakes of an exam program



MinorModerateMajor

End Section



### 4.4 Bookmark: Minor Risk






Minor

## 4.5 Minor Risks



Minor Risks

- Disruptive exam experience during test delivery (e.g., software bug requiring a restart); where there is no loss of time or response data and only one or a small group of examinees are impacted




## 4.6 Minor Risks





Minor Risks

- Incorrect non-score related information (e.g., an examinee's name or something about their demographic information)




## 4.7 Minor Risks






### Minor Risks

- Incorrect non-examinee specific info (e.g., typo/ omission/inaccuracy in score interpretation materials)



## 4.8 Bookend: Minor Risk







This is the end of this part.

Levels of Risk

#### 4.9 Bookmark: Moderate Risk




#### 4.10 Moderate Risks




Moderate Risks

- Error in non-primary inference score results (e.g., subscores, SEM, incorrect response feedback)

$$2+2=5$$




#### 4.11 Moderate Risks





Moderate Risks

- An item(s) on a form have been used too often and may contribute to content overexposure




#### 4.12 Moderate Risks

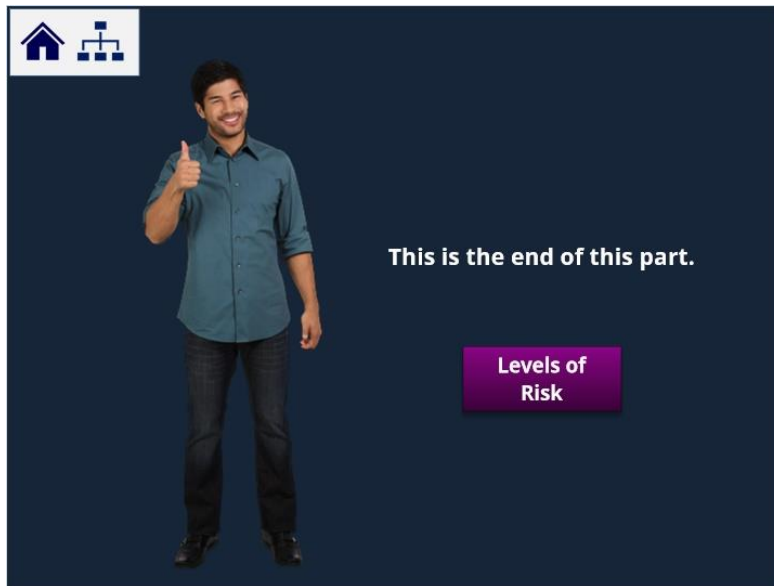


Moderate Risks

- Sample size limitations influence psychometric procedures (e.g., equating, standard setting, norm group definitions)





#### 4.13 Bookend: Moderate Risk



#### 4.14 Bookmark: Major Risk




#### 4.15 Major Risks



Major Risks

- Error in calculating scores related to primary inference (e.g., total score, classification decision)



#### 4.16 Major Risks



Major Risks

- Reliable information (data analysis, video capture, anonymous tip) suggests an examinee(s) cheated



#### 4.17 Major Risks






### Major Risks

- An upgrade to the computer-based testing hardware causes items with images to display incorrectly.



#### 4.18 Bookend: Major Risk

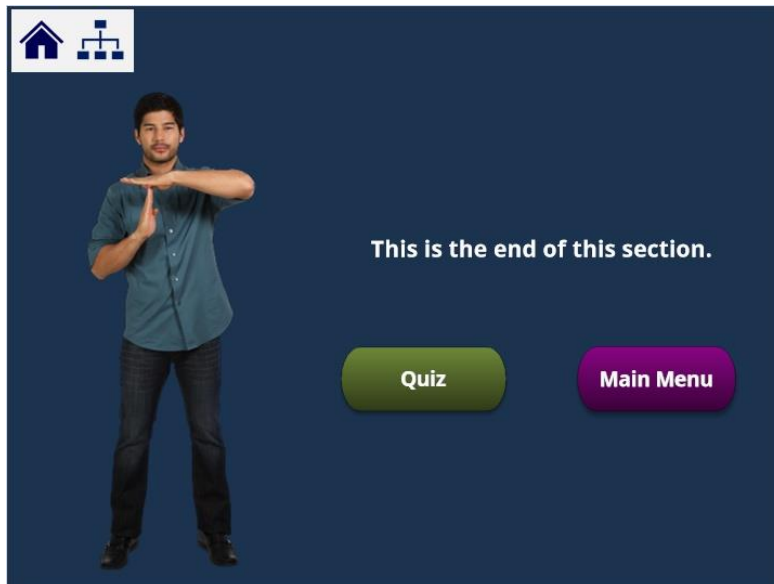




This is the end of this part.

Levels of Risk

#### ***4.19 Bookend: Risk***





## **5. Considerations and Actions**


### ***5.1 Considerations & Action Beginning***



## 5.2 Learning Objectives: Considerations



### Learning Objectives




1. Explain how each role might approach investigating an assessment challenge
2. Identify different statistical and business considerations
3. Understand how the considerations and actions as a result of the issue may differ based on the level of risk

## 5.3 Investigate Issue





### Investigate Issue

- Role Involvement:
  - Program Manager
  - Customer Support
  - Psychometrics
  - IT
  - Test Development





## ***5.4 Investigate Issue***



### **Considerations for Defining Scope**

- What is the level of risk?
- What aspects of the assessment lifecycle have been impacted?
- Who needs to do what?

## ***5.5 Statistical Considerations***





### **Statistical Considerations**

- How many examinees/items were affected?
- Which parts of the test were impacted?
- Can score information of comparable precision and validity be created?
- To what extent are score inferences compromised?



## 5.6 Business Considerations



### Business Considerations

- How many examinees/items were impacted?
- Existing policy decisions and processes
- Resources for incident management, contractual responsibility for costs
- Perception and brand risk

## 5.7 Example



**Example:** A disruption occurred while examinees were completing an assessment

**Investigation by Role:**



- Customer Support
- Information Technology
- Test Development
- Psychometrics
- Program Manager



Minor Risk      Moderate Risk      Major Risk

End Section


## 5.8 Minor Risk



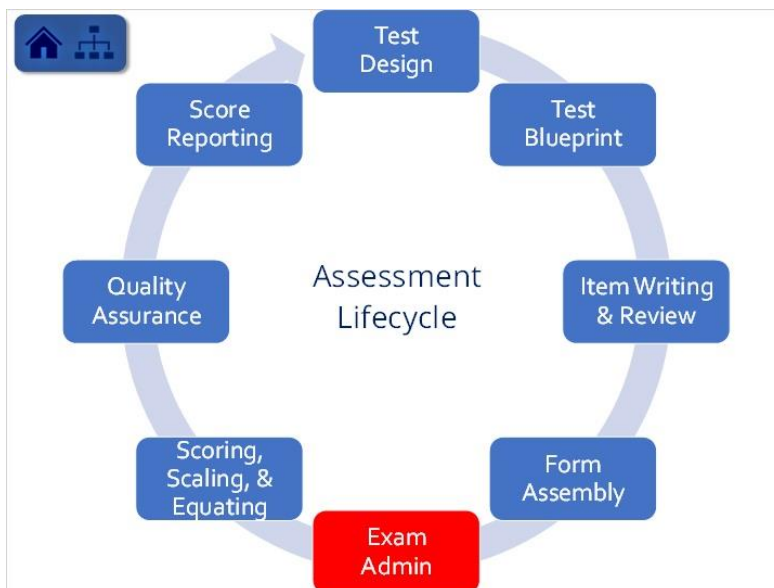
Minor Risk

- A bug with the software exam driver affects extremely few examinees resulting in a restart of their exam session. The examinees were able to continue where they left off and didn't lose any time or response data.



Minor



## 5.9 Lifecycle Example




## 5.10 How to Proceed





### Considerations and Actions

- Program manager communicates with the client
- Customer support manage examinee inquiries
- No action is warranted; the examinees response data can be scored and reported similar to the rest of the cohort.



Back to Example


## 5.11 Moderate Risk



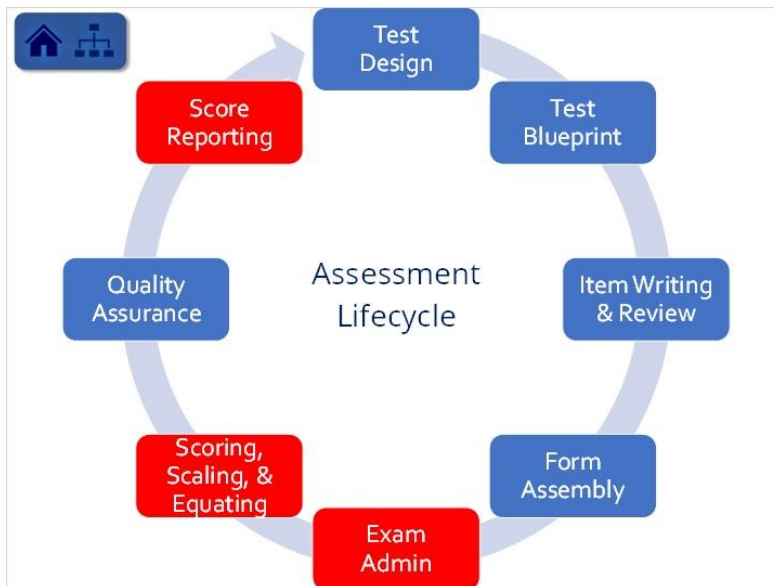
### Moderate Risk

- A bug with the software exam driver affects more than a few examinees and causes them to have 1 missed item that's scored as incorrect. The overall score impact is mostly trivial, but can be more concerning depending on performance relative to the passing threshold


Moderate



## 5.12 Lifecycle Example



## 5.13 How to Proceed



**Considerations and Actions**

Same as with a minor risk plus:

- Psychometrics, with input from other roles, investigate if any examinees performed near the passing score in which the missed item could have resulted in a different outcome
- Depending on the investigation, program manager will also consider resources for incident management, costs for retakes


[Back to Example](#)

### 5.14 Major Risk



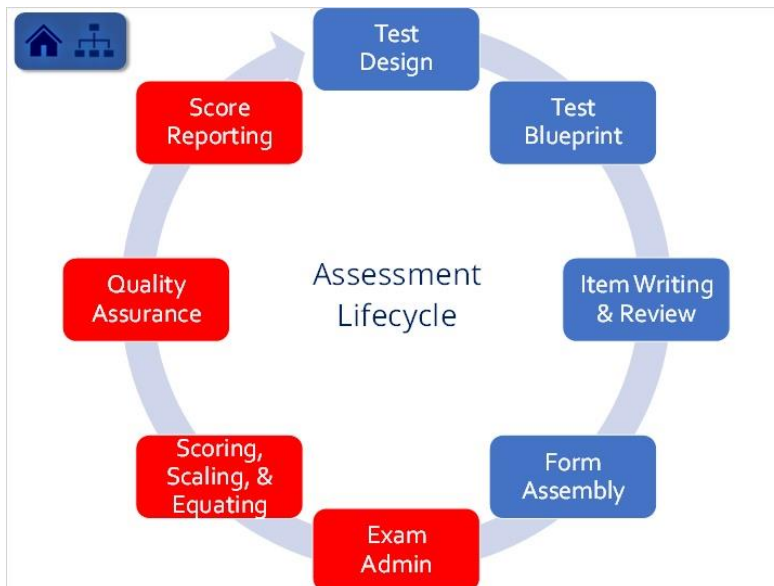
Major Risk

- A bug with the software exam driver affects a substantial number of examinees and causes them to miss a significant amount of test content.





Major

### 5.15 Lifecycle Example



## 5.16 How to Proceed





### Considerations and Actions

Same as with a moderate risk plus:

- Policy rules for defining content (e.g., how much test) and statistical (confidence in score/classification) thresholds
- If appropriate, apply statistical methods to calculate and report an examinees score/decision based on the unaffected portion of the test
- Client still the ultimate decision maker

Back to Example

## 5.17 Bookend: Considerations



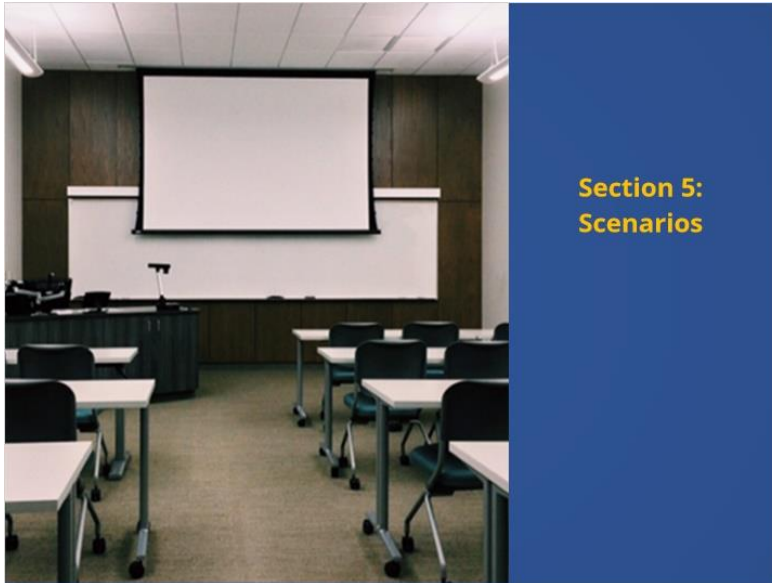


This is the end of this section.

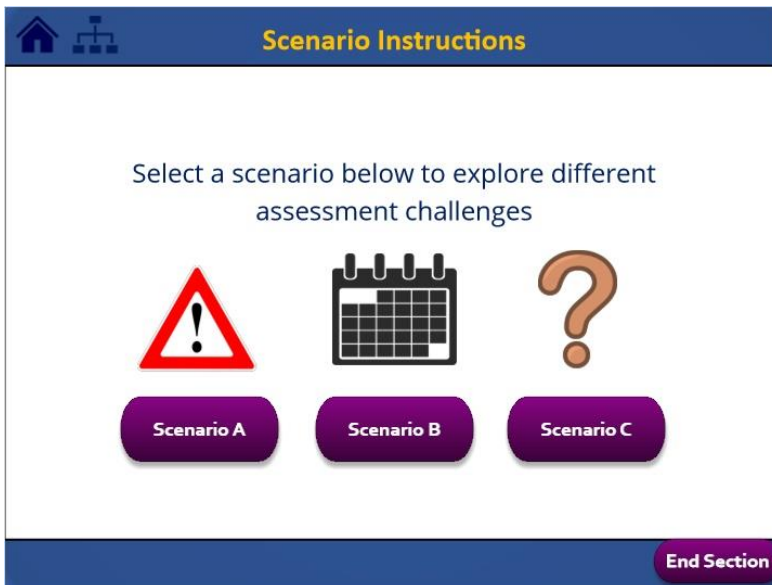
Main Menu

## 6. Scenarios

### 6.1 Scenarios





### 6.2 Scenario Intro





### 6.3 Scenario A



Scenario A

Read the following scenario and click on the buttons to learn more:

A testing program receives a message from the proctor that there may have been a cheating incident.

Roles/  
Investigate



Risk Levels

Assessment  
Lifecycle

Considerations  
and Actions

Back to Scenarios

### 6.4 Scenario A Risks




Risks

What would make this a minor, moderate, or major risk?

Moderate

Minor

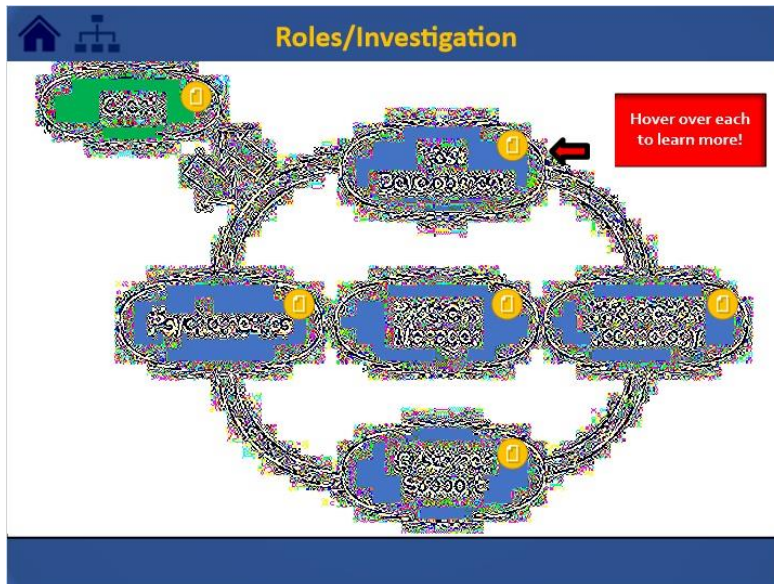
Major



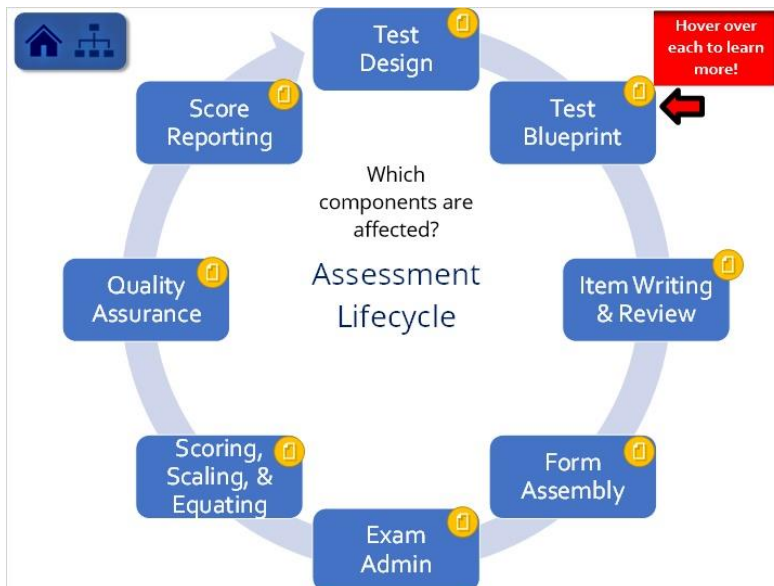
RISK

Hover over each to learn more!



## 6.5 Scenario A: Roles



## 6.6 Scenario A Cycle



## 6.7 Scenario A Considerations





Considerations and Actions

Questions that may be considered:

- Does any of the live material need to be replaced/ removed from operational scoring?
- How were security protocols breached (e.g., pre-knowledge, stealing content, answer copying)
- Do exam administrative procedures need to be adjusted?
- Can score information for suspected examinee(s) be reported?
- What is the probability of cheating?
- What other information does the client need to determine appropriate consequences

## 6.8 Scenario B



Scenario B

*Read the following scenario and click on the buttons to learn more:*

Six months before examinees register for the test, the client informs the testing program that they are relaxing eligibility requirements so that professionals from related fields may also obtain the credential.

Roles/  
Investigate

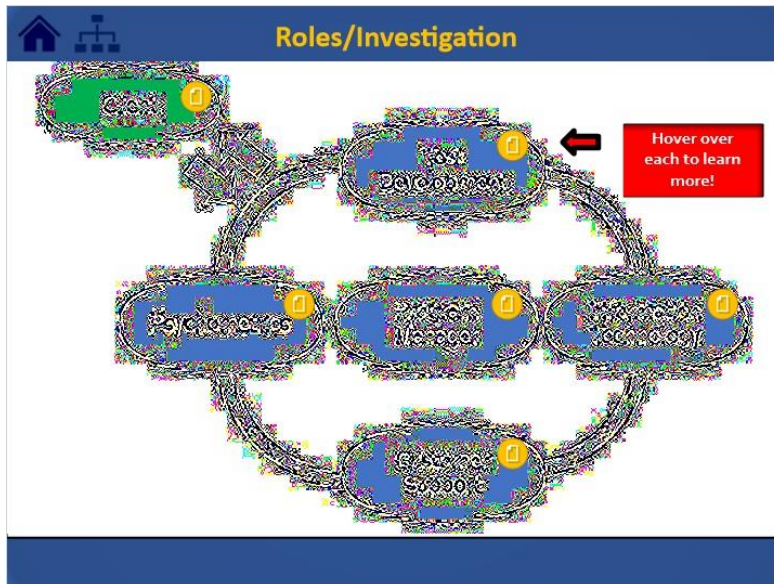
Risk Levels

Assessment  
Lifecycle

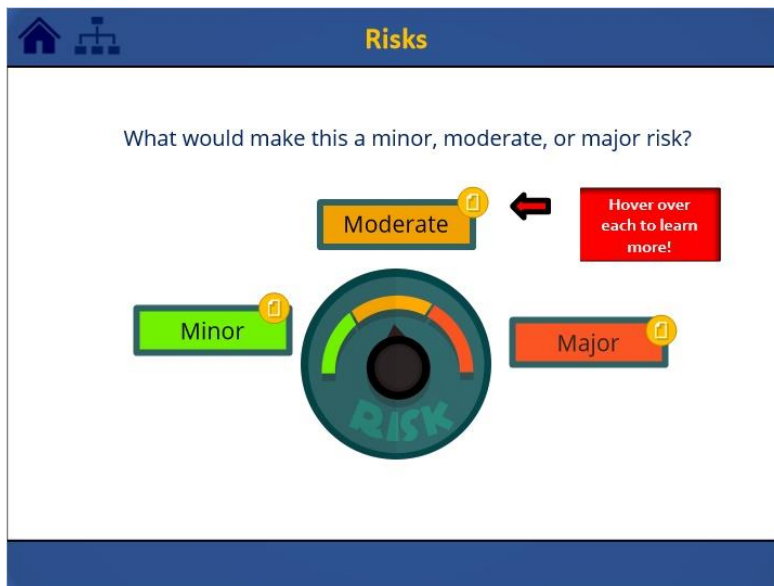
Considerations  
and Actions

Back to Scenarios

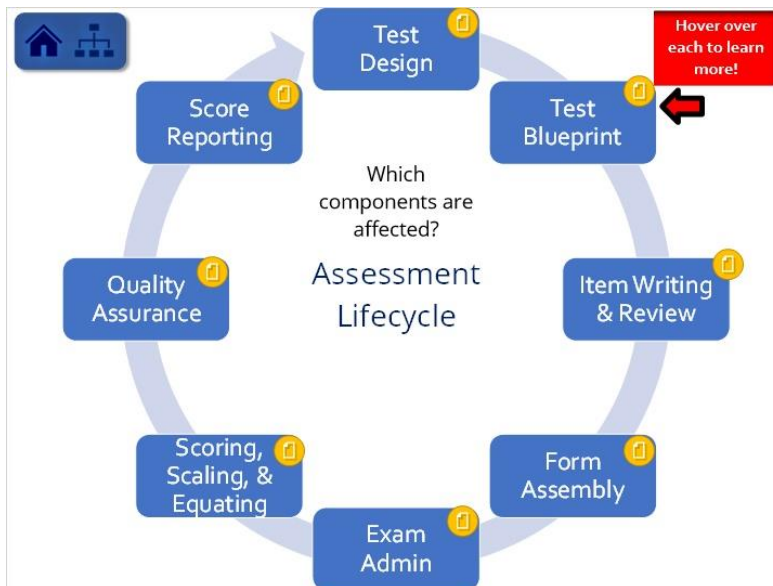
## 6.9 Scenario B Roles




## 6.10 Scenario B Risks



## 6.11 Scenario B Cycle



## 6.12 Scenario B Considerations



**Considerations and Actions**

Questions that may be considered:

- Does any of the live material need to be replaced/removed from operational scoring?
- How will the change in examinee group composition be communicated to impacted examinees and training programs?
- Does the test blueprint need to change given the addition of professionals from related fields to the examinee population?
- Does the number of items and forms developed need to increase to decrease risk of content exposure?
- Do additional days or administrative windows need to be added to accommodate more examinees?
- Do scoring, equating, and scaling procedures need to be modified?
- Does a standard setting study need to be conducted?
- Do score reports need to change?
- How can the client avoid degradation of the credential?



## 6.13 Scenario C



### Scenario C

*Read the following scenario and click on the buttons to learn more:*

In an effort to reduce costs the client informs the testing program that the number of items on the test needs to be reduced from 200 to 150.

Roles/  
Investigate

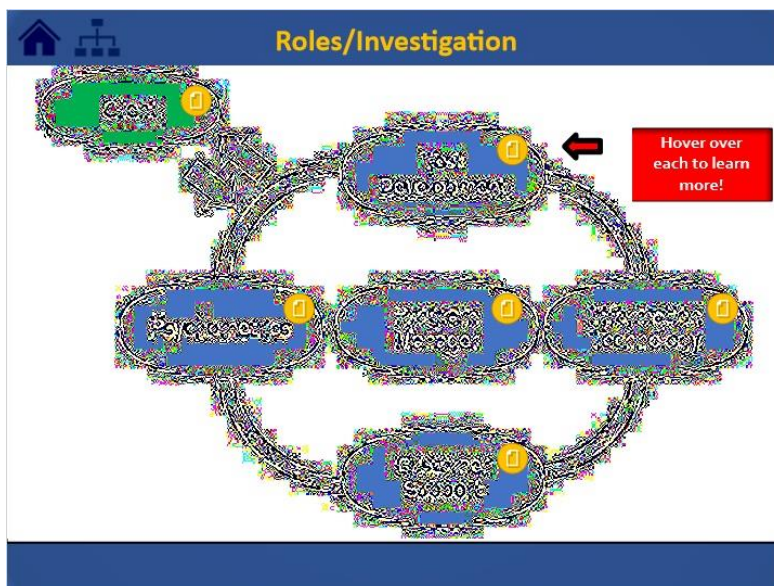
Risk Levels

Assessment  
Lifecycle

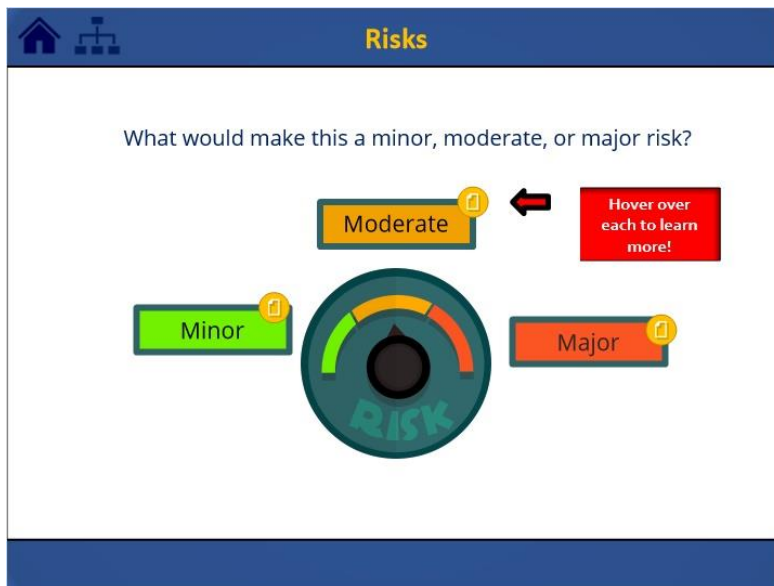
Considerations  
and Actions

Back to Scenarios

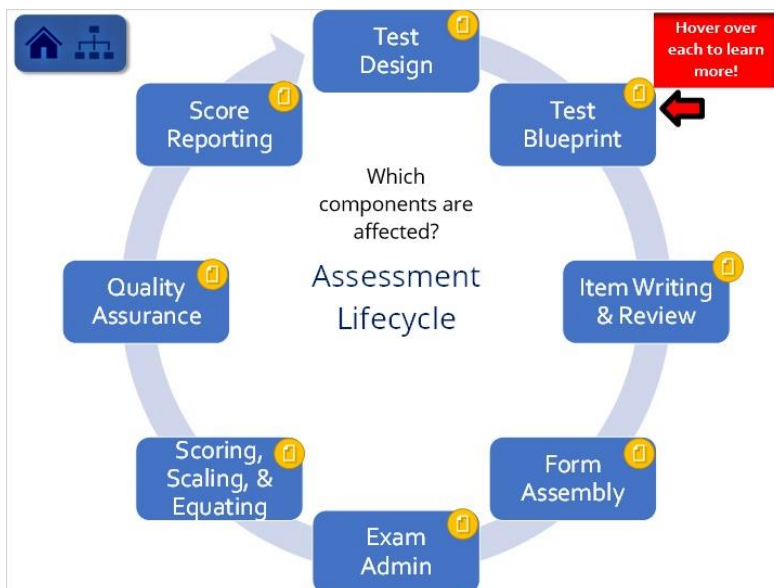
## 6.14 Scenario C Roles



## 6.15 Scenario C Risks





## 6.16 Scenario C Cycle





## 6.17 Scenario C Considerations






### Considerations and Actions

Questions that may be considered:

- Has the decision to reduce test length been communicated to examinees and other key internal and external stakeholders?
- Is the client informed of cost implications and contract modifications?
- What is the item reduction plan regarding statistical criteria, content area proportions, item format, and whom will make the decisions ?
- Does the client wish to see analyses that project the impact of item reduction on reliability and decision consistency?
- If reliability is projected to drop below what is considered best practice, what efforts will be made to recommend alternatives with the client?
- Will the change in number of items alter the construct being measured and score inferences/interpretations?
- Will testing time be decreased? Has that decision been communicated with key staff, vendors, and stakeholders?
- Will subscore reporting change?
- Will the score scale change?
- Will a standard setting study be conducted?

## 6.18 Bookend: Scenarios





This is the end of this section.

[Main Menu](#)