How Do You Know It’s a Good Test?

Mr. Joe Fish
Questionmark
How do you know it’s a good test?

Joe Fish
Product Director, Questionmark
Questionmark OnDemand for Government

Cloud-based Assessment Management System for US Government/Defense

FedRAMP

Security package now available

www.questionmark.com/trust/FedRAMP
What can we learn from assessments?

Did they like it?

Did they learn it?

Can they do it?

Can we prove it?
Types of Assessments

- Tests & Exams
- Job Task Analysis
- Quizzes
- Observational
- Surveys
Assessment results help us in many ways...

1. Make informed, defensible personnel decisions
2. Hire and promote the right people
3. Align competencies with organizational objectives
4. Target learning investments accurately
5. Ensure operational readiness
When the stakes are high, assessment results must be **trustworthy** and **defensible**
Keys to getting trustworthy results....

Assessments need to be
- Valid
- Reliable
- Secure

Sound practices for developing, delivering and analysis

Technologies to support sound assessment practices
Reliability & Validity

**Reliable:**
- Dependable, repeatable, consistent

**Valid:**
- Measures appropriate knowledge and skills

- Reliable but not Valid
- Not Reliable, not Valid
- Reliable and Valid
Steps to getting trustworthy assessment results

- Planning assessment
- Authoring items
- Assembling assessment
- Pilot and review
- Delivery
- Analyze results

Security
Validate assessment content using Job Task Analysis

Survey SMEs about key tasks: How Difficult? How Important? How Frequent? How Critical?

<table>
<thead>
<tr>
<th>Activity</th>
<th>Do Task</th>
<th>Supervise Task</th>
<th>N/A</th>
<th>Very Easy</th>
<th>Easy</th>
<th>Neither Easy or Difficult</th>
<th>Difficult</th>
<th>Disaster Management</th>
</tr>
</thead>
<tbody>
<tr>
<td>Administering medication</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td></td>
</tr>
<tr>
<td>Assessing patients</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td></td>
</tr>
<tr>
<td>Assisting patient</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td></td>
</tr>
<tr>
<td>Communicating with family members</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td></td>
</tr>
<tr>
<td>Cleaning surgical area</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td></td>
</tr>
<tr>
<td>Serving patients</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td></td>
</tr>
</tbody>
</table>

2 of 3
What is your role in the organization?

3 of 3
Answer questions about nursing.
Analyze JTA results to “blue print” your test

<table>
<thead>
<tr>
<th>Electrical Skills Test - Blueprint</th>
<th>% of test</th>
<th># items</th>
<th>Skill level</th>
<th>Criticality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Section 1: AC Motors</td>
<td>45%</td>
<td>34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.a - Lorem ipsum dolor</td>
<td>21</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>1.b - Duis aute irure</td>
<td>8</td>
<td>3</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>1.c - Excepteur sint occaecat</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Section 2: High-voltage switchgear</td>
<td>25%</td>
<td>19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.a - Neque porro quisquam</td>
<td>12</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>2.b - Modi tempora incidunt</td>
<td>7</td>
<td>3</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Section 3: Synchronous Motors</td>
<td>30%</td>
<td>23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.a - Totam rem aperiam</td>
<td>11</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3.b - Eaque ipsa quae</td>
<td>12</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

How many questions for each topic?
Enable SMEs write test items collaboratively, securely

Set role-based access to items
Blueprint guides authors on item selection

- Randomly select $\xrightarrow{}$ from topic
- Include sub-topics
- Avoid previously delivered

Selected questions - drag to set question order
- 5 random question(s) from topic "Skills Test/High Voltage Switchgear" including subtopics
- 7 random question(s) from topic "Skills Test/AC Motors" including subtopics (Avoid previously delivered)
- 8 random question(s) from topic "Skills Test/Synchronous Motors" including subtopics
When assessment stakes increase, so do security requirements.
Key threats to address

- Cheating
- Content Theft
- Impersonation
Combating and Mitigating threats

Secure authentication & strong password management

Secure apps for locked-down test delivery and monitoring

Proctoring - Live Online, On-site (in person), Record-and-review
Analyzing reliability of assessment results

Are the questions too easy? Too hard? (Should I care?)

Are the results from individual questions consistent with the overall test results?
Analyze Item Statistics: Difficulty Index

- Measures % correct responses
  - Expressed as number: 0.0 - 1.0
    - Higher = Easier
    - Lower = Harder

- Flags possible problems:
  - Too high: Too easy!
  - Too low: Too difficult?
  - Confusing/ambiguous? Miscoded?
Analyze Item Statistics: **Correlation Discrimination**

- Measures correlation between item score and test score
- Are the item results consistent with assessment results?
- Is this item a good/poor predictor of overall assessment results?
Assessment Management System

Reporting to analyze results and inform stakeholders

Collaborative, secure authoring of surveys, quizzes, tests, and exams

Blended delivery to support a broad range of devices and security needs

Enables and supports sound assessment practices
Learn more...

Come and visit us at our booth!