questionmark

Compass Consultants, LLC

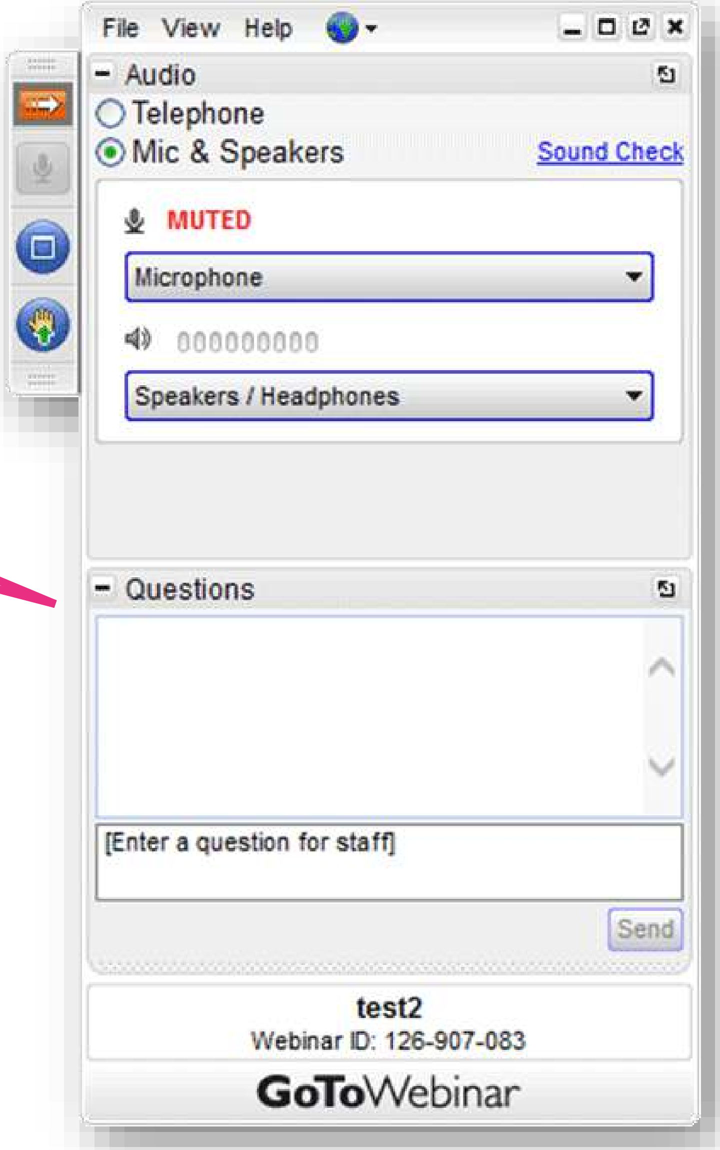
Test-Item Database Design – Your Key to Fairness

Jim Parry, M.Ed., CPT, Compass Consultants, LLC

August 26, 2021

This Photo by Unknown Author is licensed under CC BY-SA

To ask questions, use the “Questions” feature



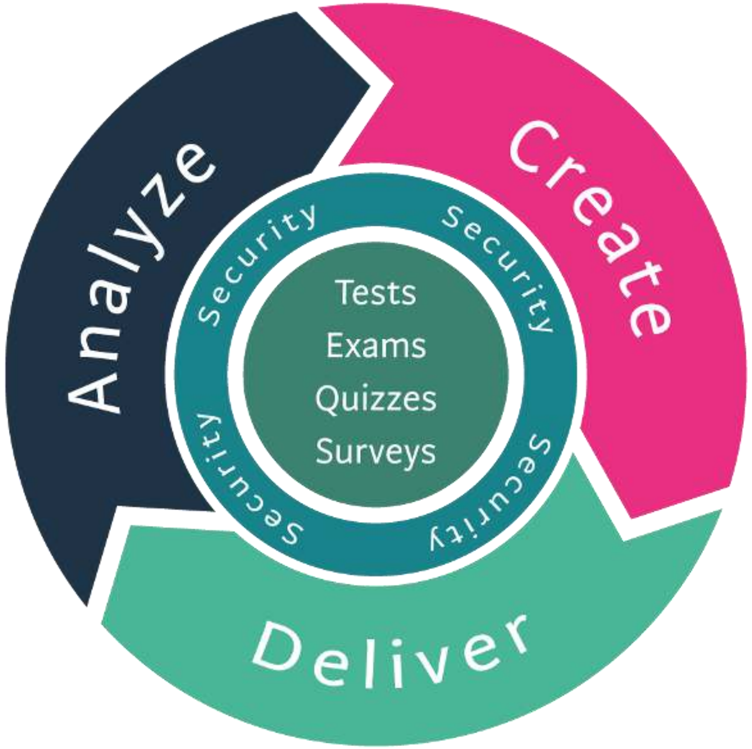
Watch for an email after the webinar:

- Download slides (PDF)
- View a recording
- Answer a survey

About Questionmark

Background

- Founded in 1988
- Assessment solutions to measure knowledge, skills, abilities and attitudes securely for certification, regulatory compliance, workforce learning, sales-force readiness and higher education
- ISO/IEC 27001 Certified (Learn more: www.questionmark.com/trust)



- *Questionmark OnDemand*
- *Questionmark OnDemand for Government*
- *Questionmark OnPremise*

Today's Presenter

Jim Parry, M.Ed., CPT, Compass Consultants, LLC

- Owner and Chief Executive Manager of Compass Consultants, LLC
- Over 40 years' experience in course design, development, presentation and assessment design and analysis
- Holds a Master of Education degree from the University of West Florida and is a Certified Performance Technologist (CPT), awarded by the International Society of Performance Improvement (ISPI)
- Has been presenter of pre-conference workshops and educational sessions at various professional conferences for many years
- Internationally recognized consultant providing services concerning test design, development, establishment of cut scores, and analysis
- Jim is a consulting partner of Questionmark

About Compass Consultants, LLC

Background

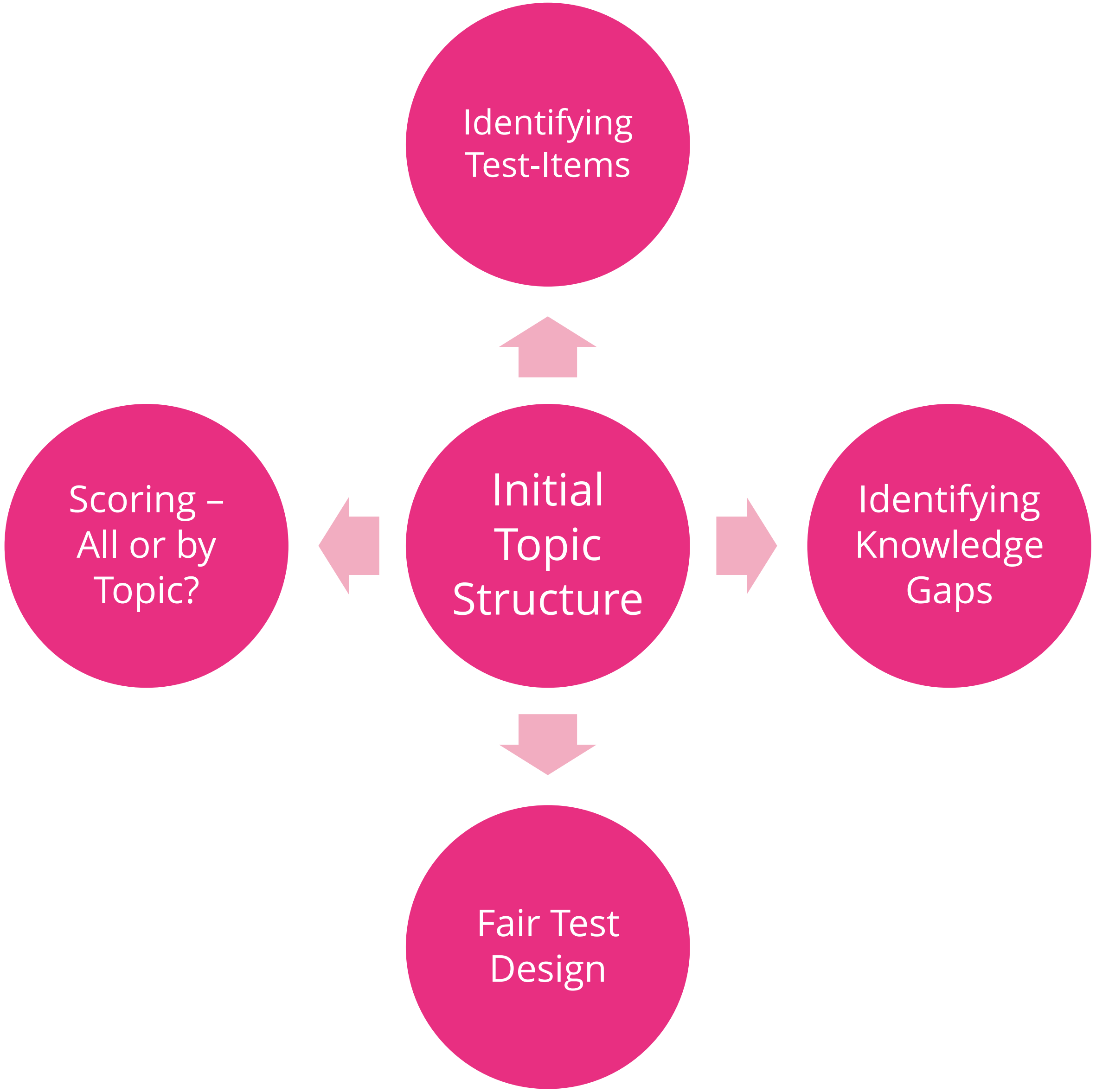
- Founded in 2010
- A leader in the application of Human Performance Technology (HPT), specializing in the design, development and presentation of training interventions and the psychometrics of test development and analysis.
- Learn more: www.gocompassconsultants.com

Compass Consultants, LLC

Legal Disclaimer

- The presentation may include information about legal issues and legal developments. Such materials are for informational and/or educational purposes only and may not reflect the most current legal developments. These informational/educational materials are not intended, and should not be taken, as legal advice on any particular set of facts or circumstances. You should contact an attorney for advice on specific legal problems or questions.
- Information and/or software tools are provided "as is" without any express or implied warranty of any kind including warranties of merchantability, noninfringement of intellectual property, or fitness for any particular purpose. In no event shall Compass Consultants, LLC., or its employees, contractors, sub-contractors, agents, officers or attorneys be liable for any damages whatsoever (including, without limitation, damages for loss of profits, business interruption, loss of information) arising out of the use of or inability to use the information, even if Compass Consultants, LLC has been advised of the possibility of such damage.

Agenda


Identifying Test-Items
Scoring – All or by Topic?
Initial Topic Structure
Identifying Knowledge Gaps
Fair Test Design

question mark

Compass Consultants, LLC

Setting Up the Topic Structure

How Low Should You Go?

Quick Poll ☑

How does your organization arrange the test-item database?

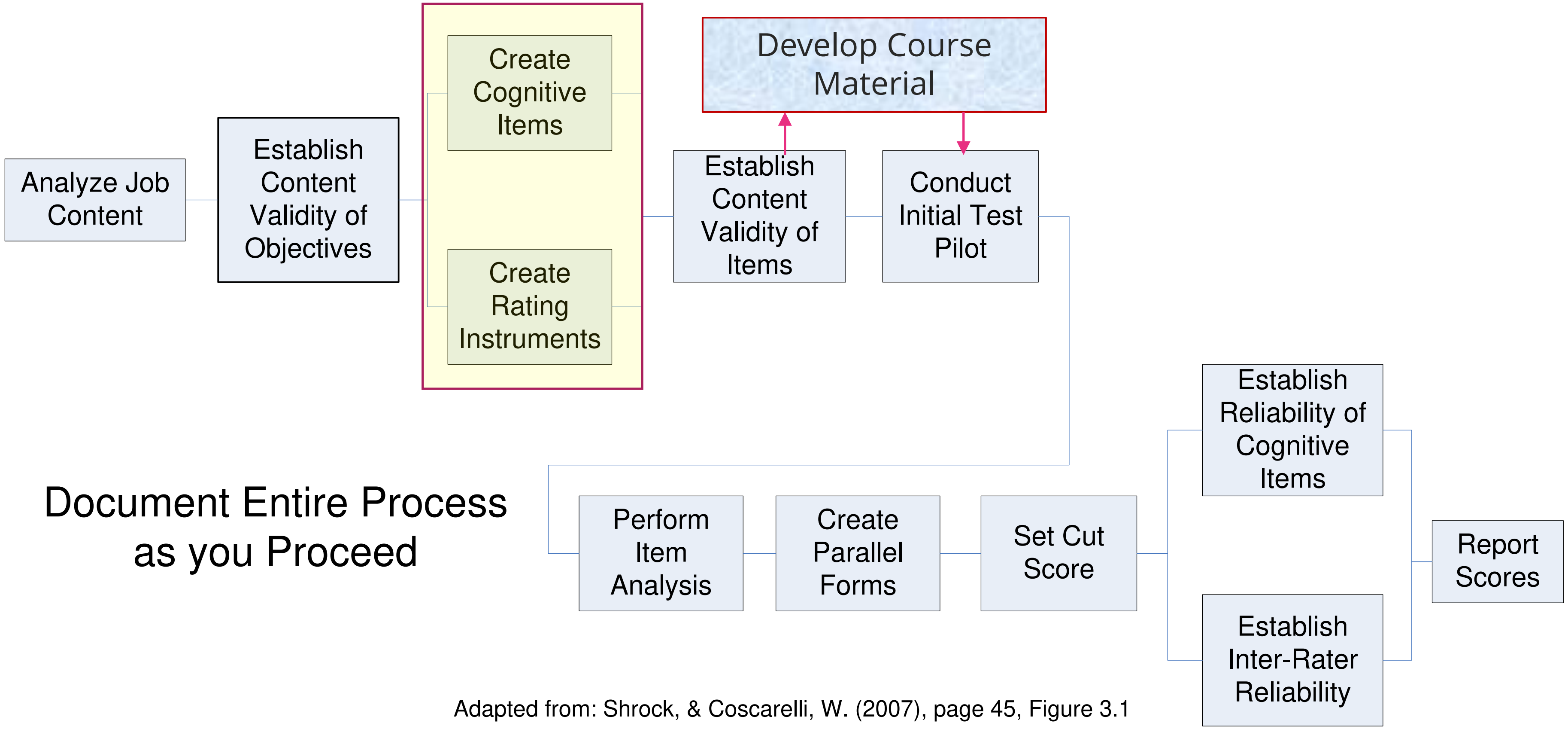
A. All test-items (questions) are stored under one topic

B. Test-items are stored in separate topics and sub-topics under a main topic

C. We have not started building our database yet

Where it Begins

Designing Criterion-Referenced Tests



Analyze Job Content

Establish Content Validity of Objectives

Create Cognitive Items

Create Rating Instruments

Develop Course Material

Establish Content Validity of Items

Conduct Initial Test Pilot

Perform Item Analysis

Create Parallel Forms

Set Cut Score

Establish Reliability of Cognitive Items

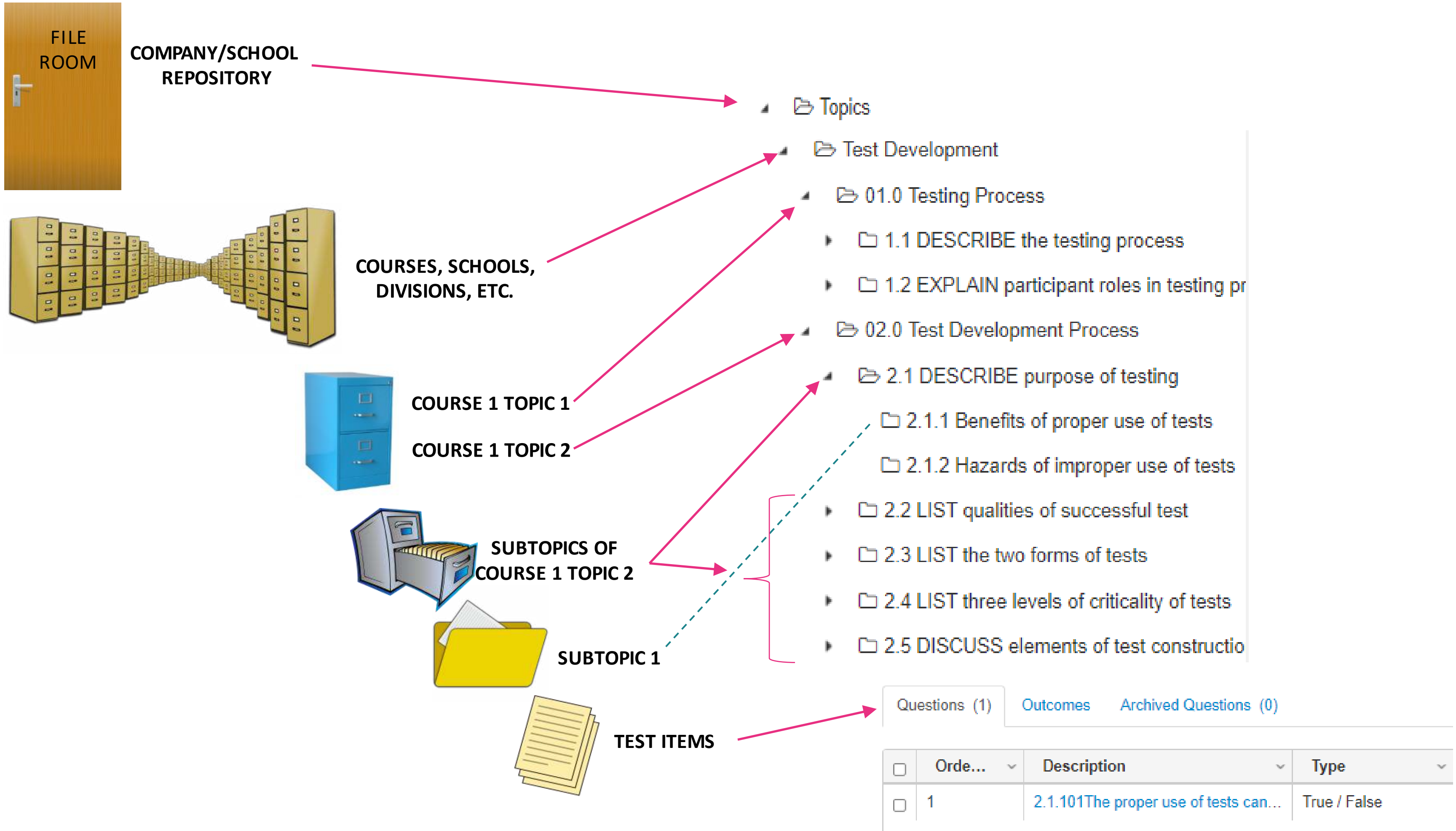
Establish Inter-Rater Reliability

Report Scores

Document Entire Process as you Proceed

Adapted from: Shrock, & Coscarelli, W. (2007), page 45, Figure 3.1

Match Topic Structure to Curriculum Outline



COMPANY/SCHOOL REPOSITORY

COURSES, SCHOOLS, DIVISIONS, ETC.

COURSE 1 TOPIC 1

COURSE 1 TOPIC 2

SUBTOPICS OF COURSE 1 TOPIC 2

SUBTOPIC 1

TEST ITEMS

- Topics
  - Test Development
    - 01.0 Testing Process
      - 1.1 DESCRIBE the testing process
      - 1.2 EXPLAIN participant roles in testing pr
    - 02.0 Test Development Process
      - 2.1 DESCRIBE purpose of testing
        - 2.1.1 Benefits of proper use of tests
        - 2.1.2 Hazards of improper use of tests
      - 2.2 LIST qualities of successful test
      - 2.3 LIST the two forms of tests
      - 2.4 LIST three levels of criticality of tests
      - 2.5 DISCUSS elements of test constructio

Questions (1) | Outcomes | Archived Questions (0)

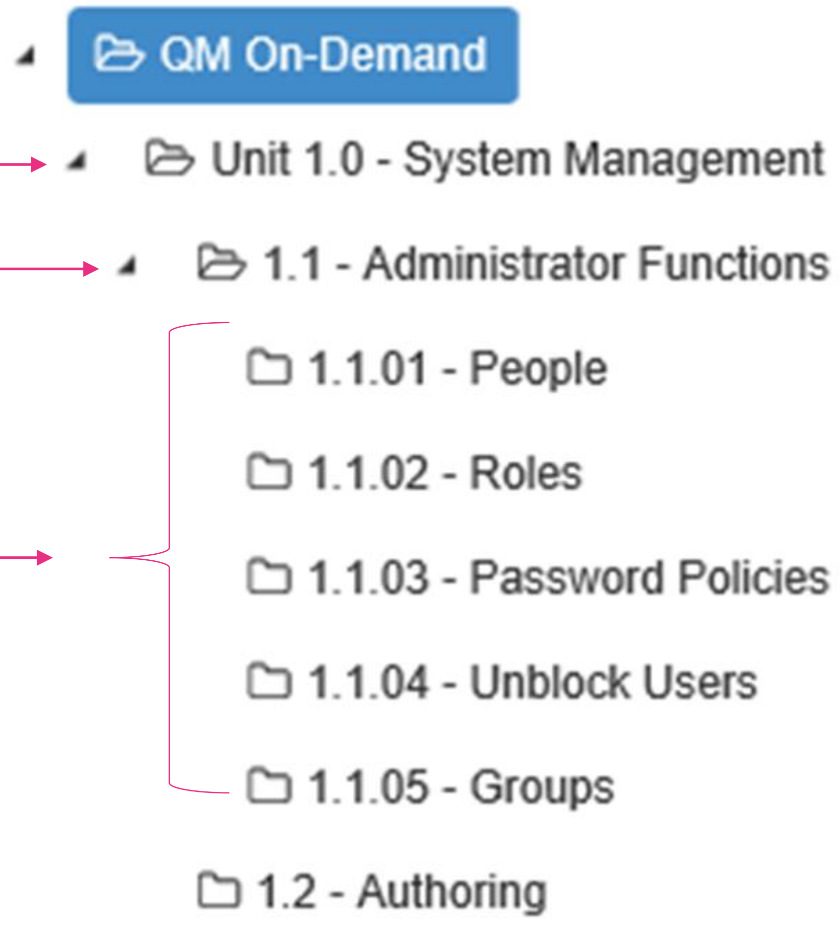
□	Orde...	Description	Type
□	1	2.1.101The proper use of tests can...	True / False

Plan Ahead

- When setting up your initial test item database think about what you want to know about the results
- Set up database by subject/topic
- Important to set up topic structure at lowest probable reporting level
- If you didn’t set up deep enough initially you can’t go back later
- Don’t worry about going too deep
- Only report to level needed

Possible Reporting Levels

- Overall “QM On-Demand” results
- By unit(s)
- By sub-topic(s) within a unit
- By sub-sub-topic(s) within a unit



- QM On-Demand
  - Unit 1.0 - System Management
    - 1.1 - Administrator Functions
      - 1.1.01 - People
      - 1.1.02 - Roles
      - 1.1.03 - Password Policies
      - 1.1.04 - Unblock Users
      - 1.1.05 - Groups
    - 1.2 - Authoring

questionmark

Compass Consultants, LLC

Assigning Question Identifiers

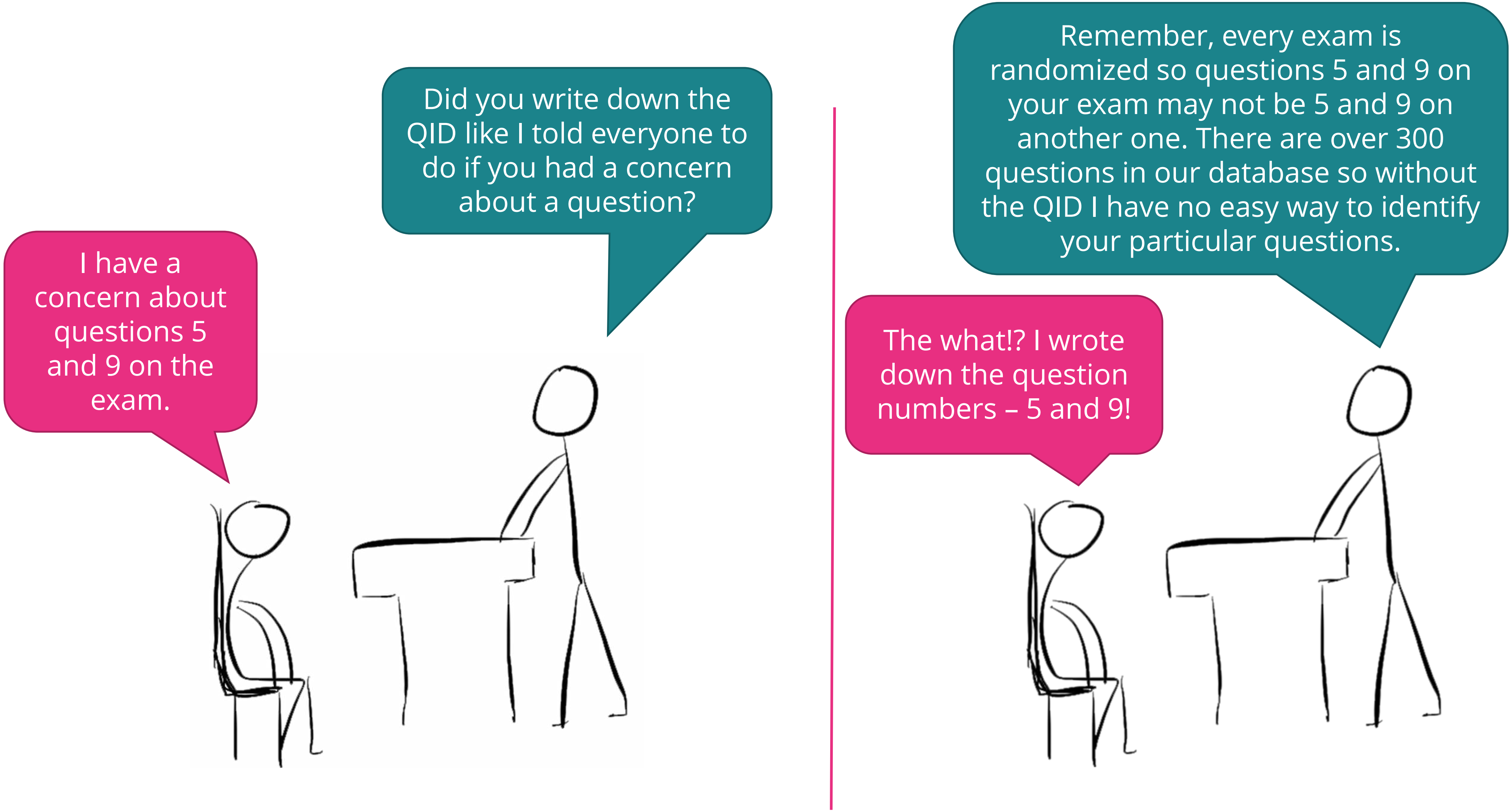
Location, Location, Location…

Quick Poll ☑

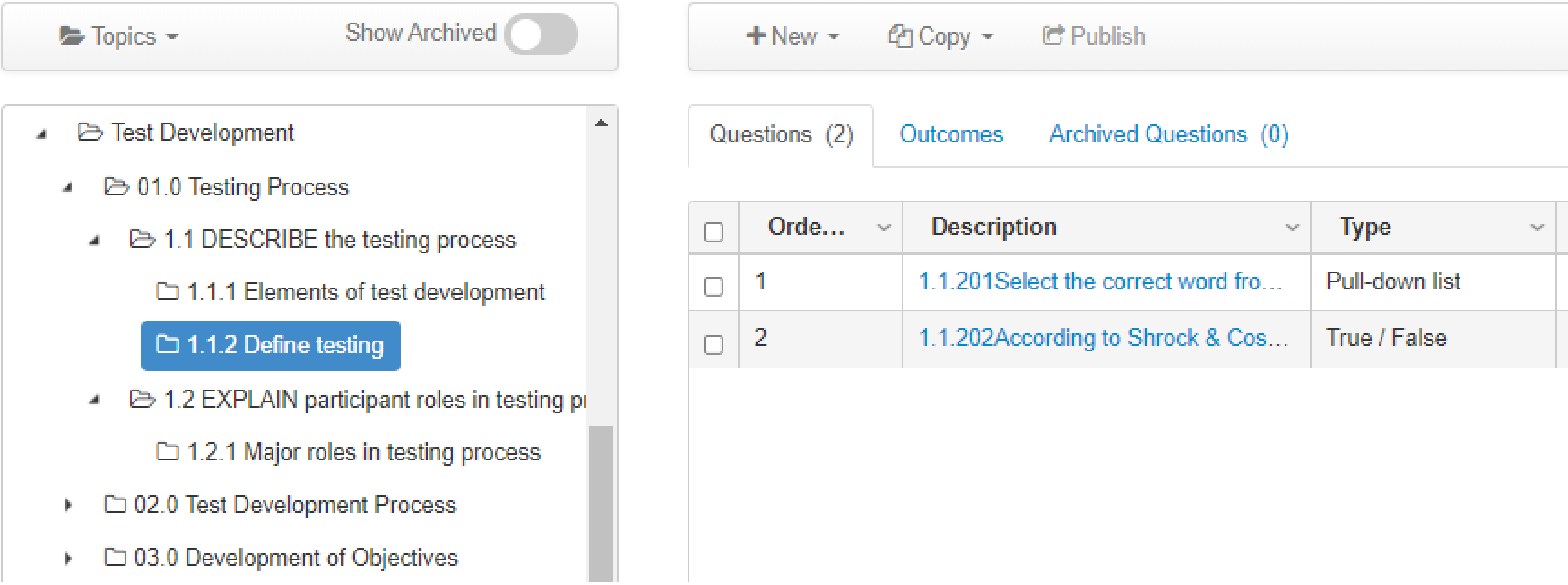
How does your organization identify test-items within the database?

A. Test-items are assigned a number to correspond to the objectives

B. Test-items are listed without numbering using the description field which is a copy of the question wording


I have a concern about questions 5 and 9 on the exam.
Did you write down the QID like I told everyone to do if you had a concern about a question?
The what!? I wrote down the question numbers – 5 and 9!
Remember, every exam is randomized so questions 5 and 9 on your exam may not be 5 and 9 on another one. There are over 300 questions in our database so without the QID I have no easy way to identify your particular questions.

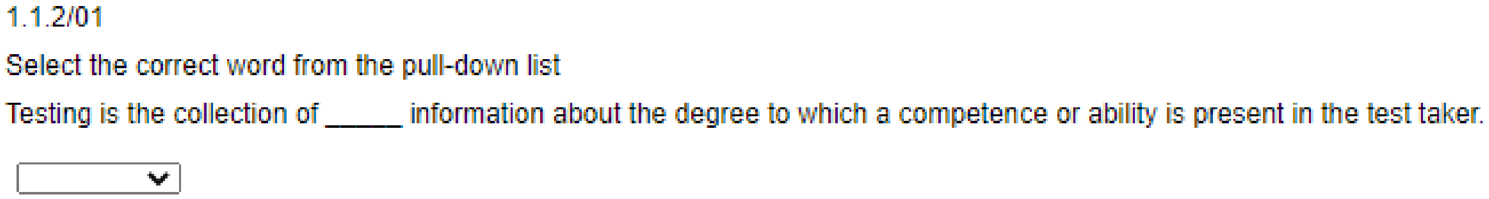
Identifying Test-Items (QID)



- Test Development
  - 01.0 Testing Process
    - 1.1 DESCRIBE the testing process
      - 1.1.1 Elements of test development
      - 1.1.2 Define testing
    - 1.2 EXPLAIN participant roles in testing pı
      - 1.2.1 Major roles in testing process
  - 02.0 Test Development Process
  - 03.0 Development of Objectives

□	Orde...	Description	Type
□	1	1.1.201Select the correct word fro...	Pull-down list
□	2	1.1.202According to Shrock & Cos...	True / False

Test-Taker sees this

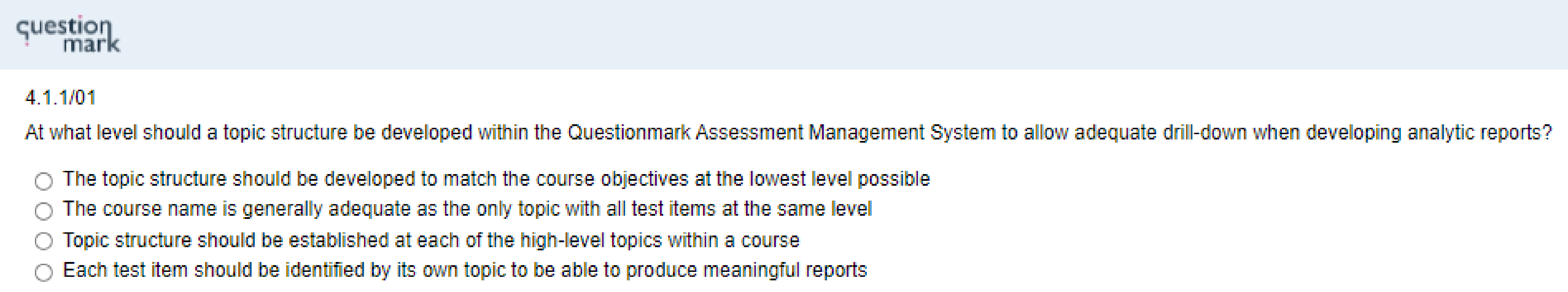


1.1.2/01

Select the correct word from the pull-down list

Testing is the collection of _____ information about the degree to which a competence or ability is present in the test taker.

Example QID



4.1.1/01

At what level should a topic structure be developed within the Questionmark Assessment Management System to allow adequate drill-down when developing analytic reports?

- ○ The topic structure should be developed to match the course objectives at the lowest level possible
- ○ The course name is generally adequate as the only topic with all test items at the same level
- ○ Topic structure should be established at each of the high-level topics within a course
- ○ Each test item should be identified by its own topic to be able to produce meaningful reports

“4.” represents the high-level topic “*Topic Structure Development*”
“4.1” identifies the Terminal Performance Objective “*DEVELOP a topic structure within the Questionmark Assessment Management System*”
“4.1.1” takes us to the first Enabling Objective **“***DISCUSS the importance of developing a topic structure to the lowest objective level*”
“4.1.1/01” indicates that this is the first test item created in the subtopic

Locating Test Items Without QIDs

- What if numerous test item stems begin the same way?
  - In which step shown below does….
  - In which step shown below does….
  - In which step shown below does….
  - In which step shown below does….
- Remember – the description shown is a copy of the first part of the stem unless it is customized

question mark

Compass Consultants, LLC

Identifying Knowledge Gaps

Where is the weakness?

Quick Poll ☑

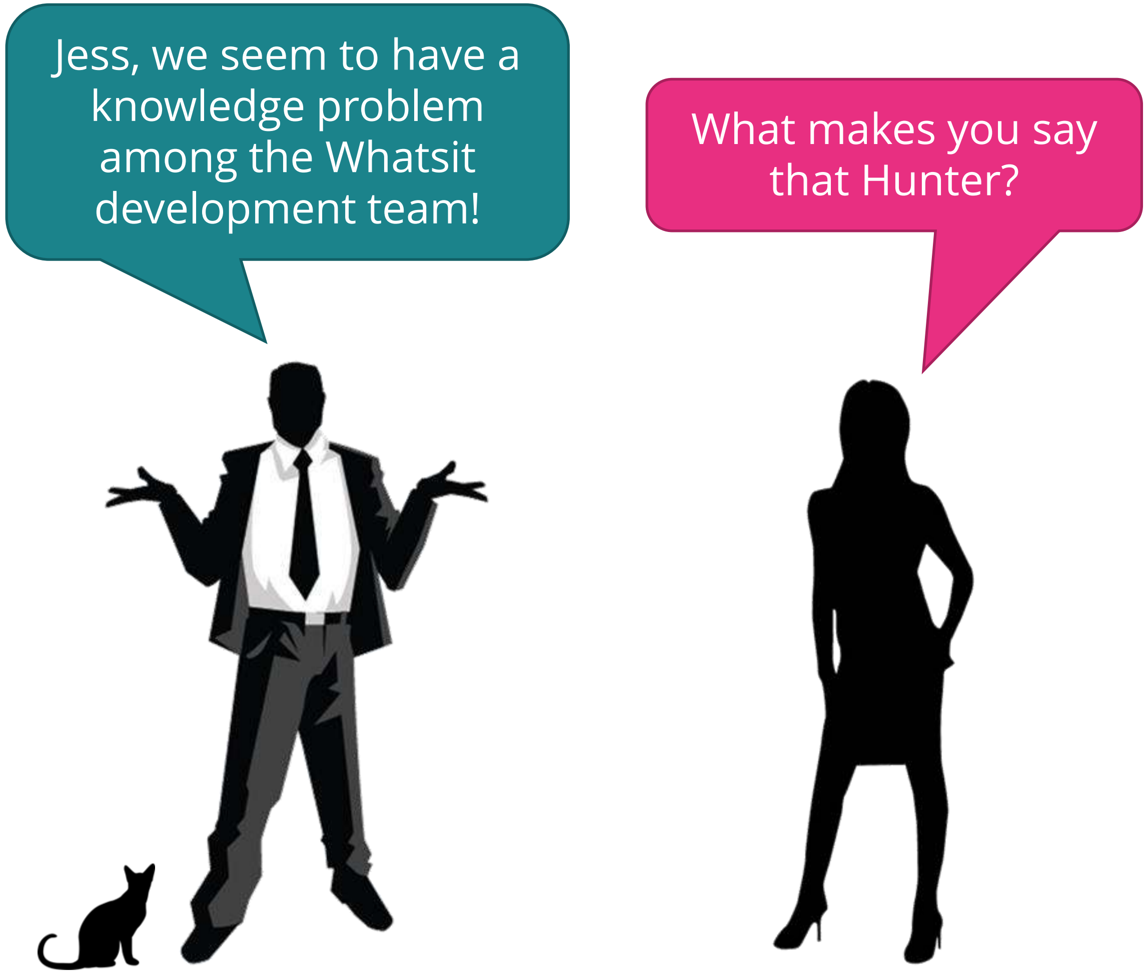
How does your organization identify knowledge gaps using test results?

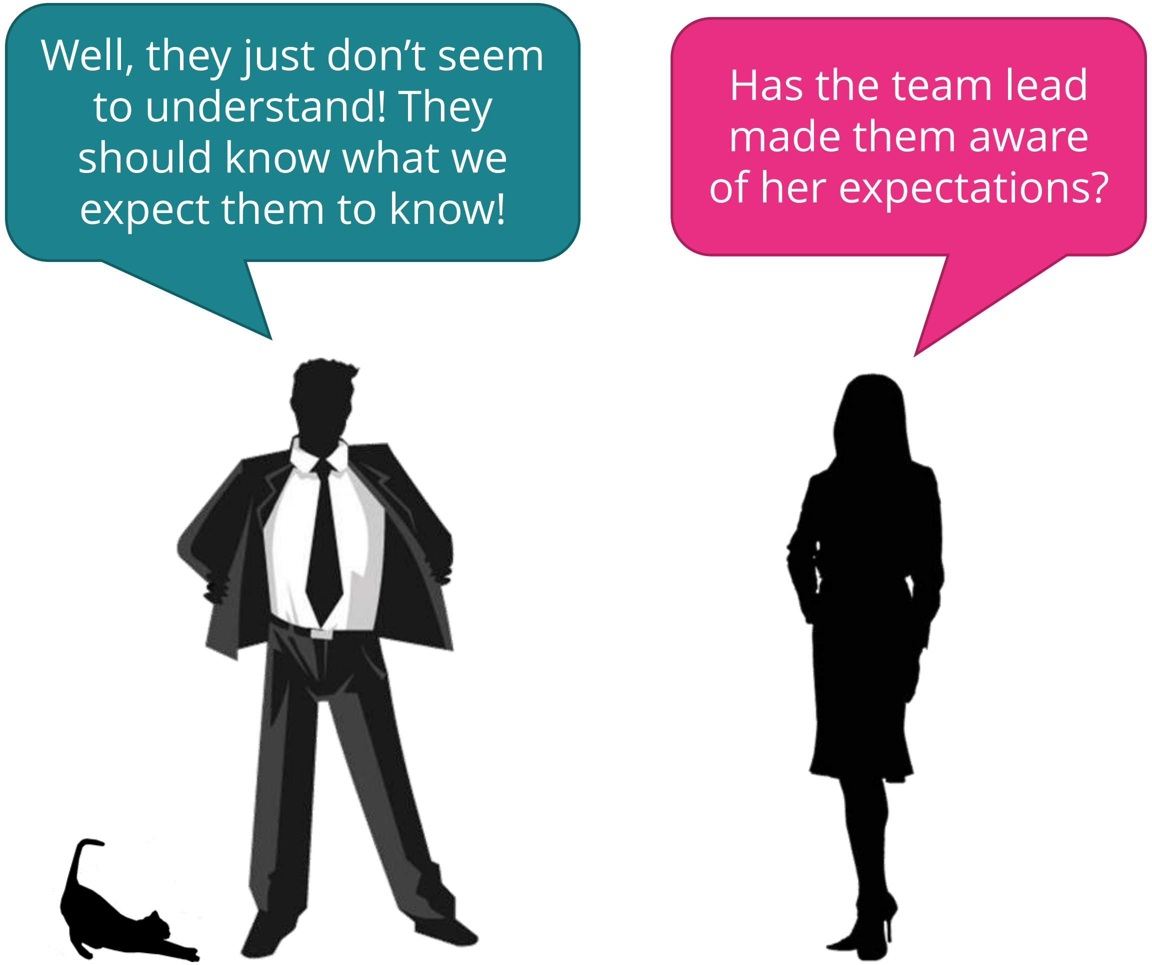
A. All tests only cover a single topic so we just look at the overall average

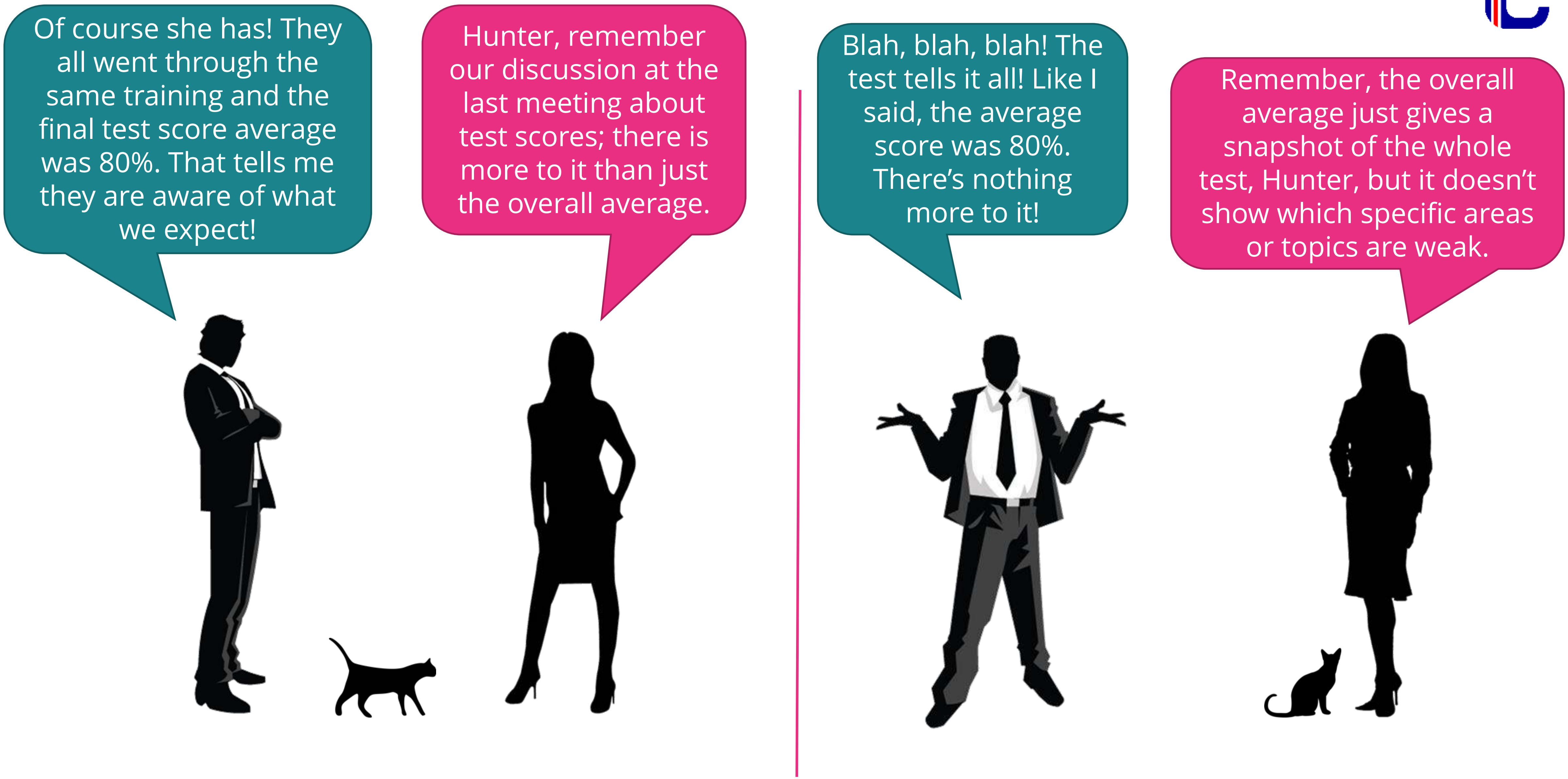
B. Even if a test covers multiple topics we only look at the overall average

C. We review the results for each topic within a test to look for weaknesses

D. What’s a knowledge gap?


Jess, we seem to have a knowledge problem among the Whatsit development team!
What makes you say that Hunter?

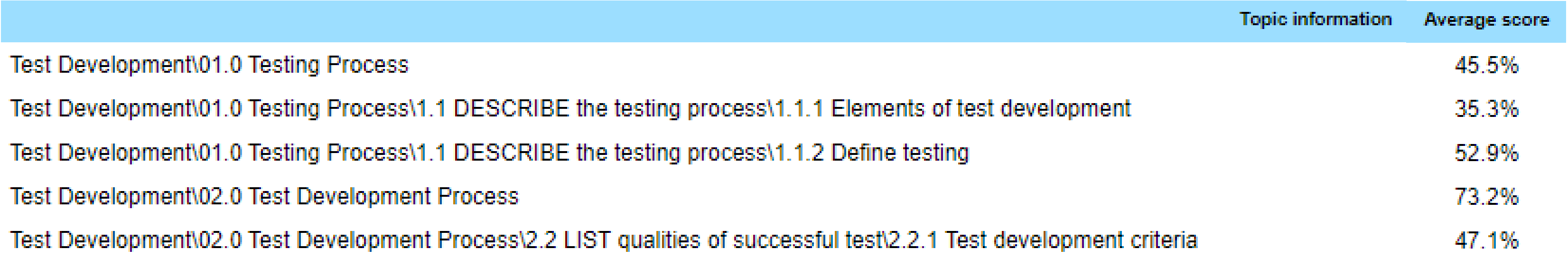

Well, they just don't seem to understand! They should know what we expect them to know!
Has the team lead made them aware of her expectations?


Of course she has! They all went through the same training and the final test score average was 80%. That tells me they are aware of what we expect!
Hunter, remember our discussion at the last meeting about test scores; there is more to it than just the overall average.
Blah, blah, blah! The test tells it all! Like I said, the average score was 80%. There’s nothing more to it!
Remember, the overall average just gives a snapshot of the whole test, Hunter, but it doesn’t show which specific areas or topics are weak.

Yeah, I do seem to remember hearing something like that. Is there a way you can see if there are any specific weak areas?
Sure thing Hunter! I will run some reports in Questionmark and see what I can come up with.
Hi Madison. Could you run the analytics on the quizzes and final exams from the Whatsit project class. Be sure to sort by topic. We need to see if there are any specific weaknesses in the training.
No problem Jess! I'll have it to you right after lunch.

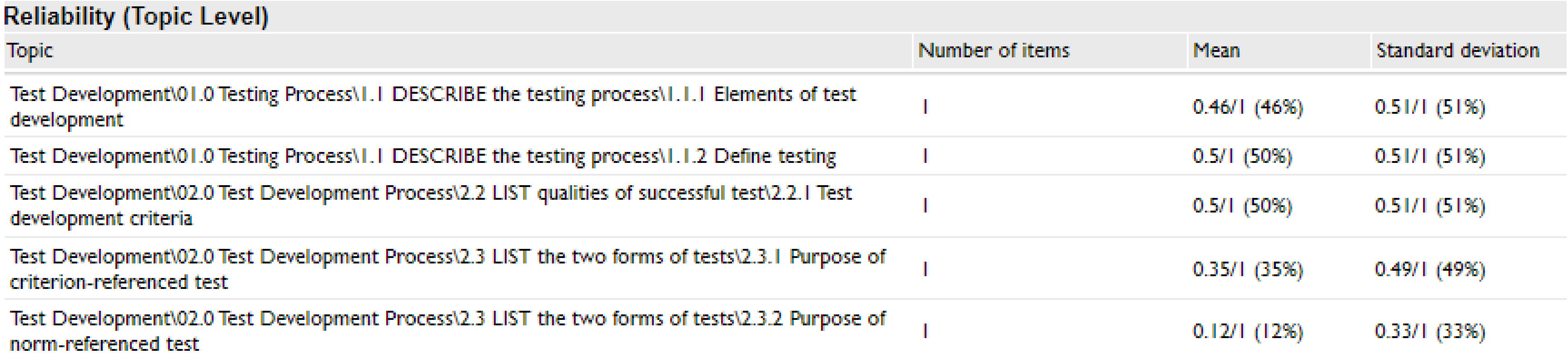
Sample Topic Reports Available in Questionmark

Assessment Overview Report



Topic information	Average score
Test Development\01.0 Testing Process	45.5%
Test Development\01.0 Testing Process\1.1 DESCRIBE the testing process\1.1.1 Elements of test development	35.3%
Test Development\01.0 Testing Process\1.1 DESCRIBE the testing process\1.1.2 Define testing	52.9%
Test Development\02.0 Test Development Process	73.2%
Test Development\02.0 Test Development Process\2.2 LIST qualities of successful test\2.2.1 Test development criteria	47.1%

Test Analysis Report

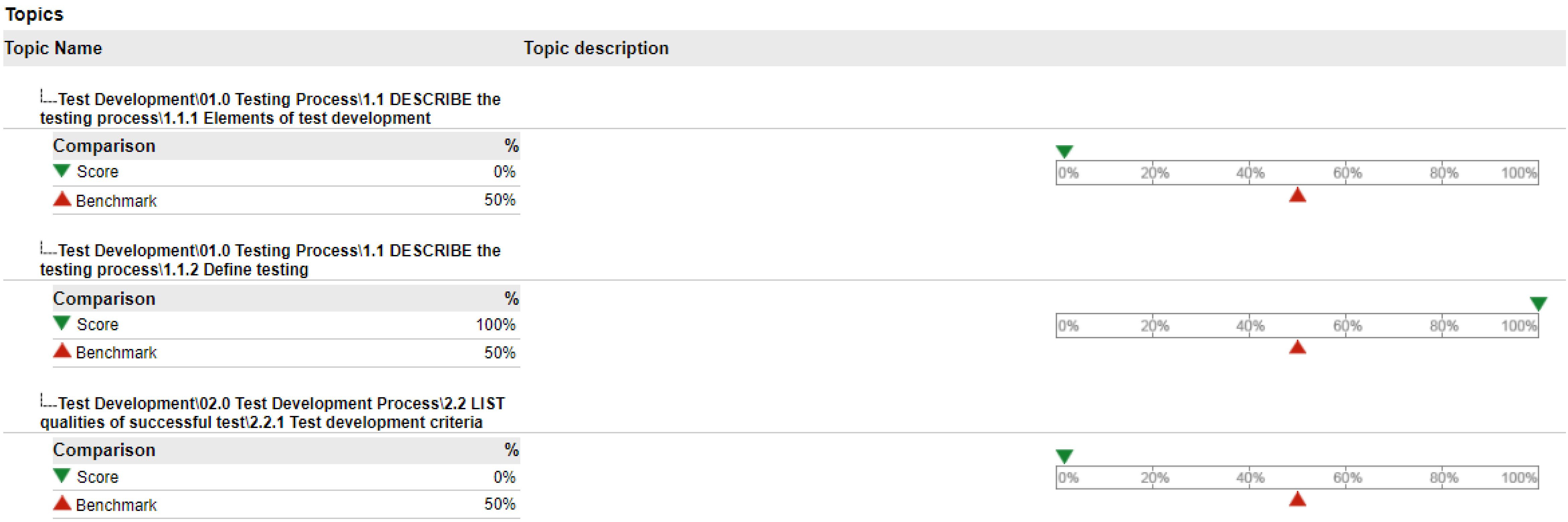


Reliability (Topic Level)

Topic	Number of items	Mean	Standard deviation
Test Development\01.0 Testing Process\1.1 DESCRIBE the testing process\1.1.1 Elements of test development	1	0.46/1 (46%)	0.51/1 (51%)
Test Development\01.0 Testing Process\1.1 DESCRIBE the testing process\1.1.2 Define testing	1	0.5/1 (50%)	0.51/1 (51%)
Test Development\02.0 Test Development Process\2.2 LIST qualities of successful test\2.2.1 Test development criteria	1	0.5/1 (50%)	0.51/1 (51%)
Test Development\02.0 Test Development Process\2.3 LIST the two forms of tests\2.3.1 Purpose of criterion-referenced test	1	0.35/1 (35%)	0.49/1 (49%)
Test Development\02.0 Test Development Process\2.3 LIST the two forms of tests\2.3.2 Purpose of norm-referenced test	1	0.12/1 (12%)	0.33/1 (33%)

Sample Topic Reports Available in Questionmark

Coaching Report



Topics

Topic Name	Topic description
Test Development\01.0 Testing Process\1.1 DESCRIBE the testing process\1.1.1 Elements of test development	

Comparison	%
▼ Score	0%
▲ Benchmark	50%

Topic Name	Topic description
Test Development\01.0 Testing Process\1.1 DESCRIBE the testing process\1.1.2 Define testing	

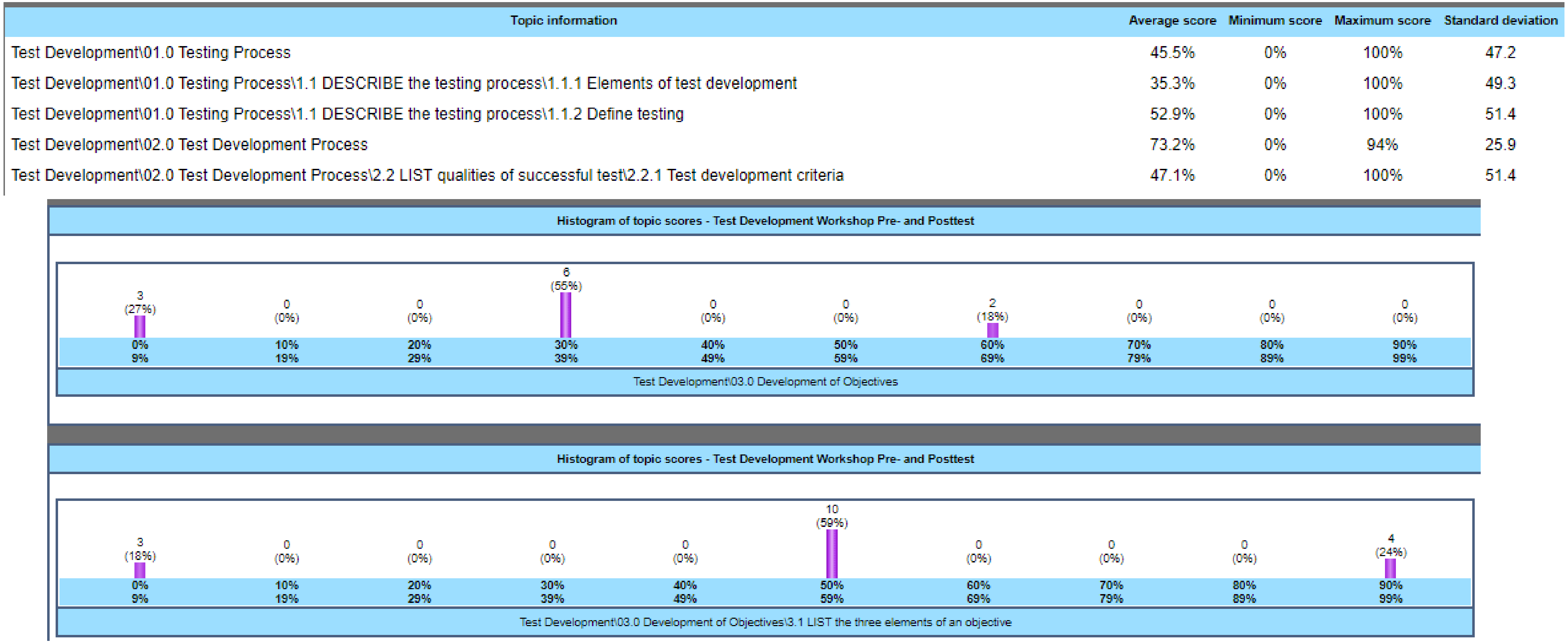
Comparison	%
▼ Score	100%
▲ Benchmark	50%

Topic Name	Topic description
Test Development\02.0 Test Development Process\2.2 LIST qualities of successful test\2.2.1 Test development criteria	

Comparison	%
▼ Score	0%
▲ Benchmark	50%

Report Manager - Design Your Own

Assessment Overview Template – Topic Performance



Topic information	Average score	Minimum score	Maximum score	Standard deviation
Test Development\01.0 Testing Process	45.5%	0%	100%	47.2
Test Development\01.0 Testing Process\1.1 DESCRIBE the testing process\1.1.1 Elements of test development	35.3%	0%	100%	49.3
Test Development\01.0 Testing Process\1.1 DESCRIBE the testing process\1.1.2 Define testing	52.9%	0%	100%	51.4
Test Development\02.0 Test Development Process	73.2%	0%	94%	25.9
Test Development\02.0 Test Development Process\2.2 LIST qualities of successful test\2.2.1 Test development criteria	47.1%	0%	100%	51.4

Histogram of topic scores - Test Development Workshop Pre- and Posttest

Test Development\03.0 Development of Objectives

Histogram of topic scores - Test Development Workshop Pre- and Posttest

Test Development\03.0 Development of Objectives\3.1 LIST the three elements of an objective

Determining a Knowledge Gap

- In order to state that there is a “gap” in knowledge, there must first be an expected level of knowledge
- If we do not identify what the minimum acceptable level of knowledge is for each part of each job, how can we identify where knowledge gaps exist?
- If an assessment covers several topics but only reports an overall score, how can a knowledge gap be identified?

False Knowledge Gap Example

Scores on Electrical Safety Assessment

Number Attaining Score	Score Range
5	100%
5	90% – 99%
15	80% - 89%
20	70% - 79%
20	60% - 69%
35	<60%

Assumptions Based on 100 Test-Takers

- Score ≥60% means satisfactory knowledge of electrical safety
- Score <60% means NOT satisfactory knowledge
- Perceived “knowledge gap” is 35%
  - 35 participants scored 60% or below

Another False Knowledge Gap Example

Scores on Electrical Safety Assessment

Number Attaining Score	Score Range
5	100%
5	90% – 99%
15	80% - 89%
20	70% - 79%
20	60% - 69%
35	<60%

Assumptions Based on 100 Test-Takers

- Minimum standard is 100%
- Score <100% means NOT satisfactory knowledge
- Perceived "knowledge gap" is 95%
    - 95 participants scored below 100%

What is on the Assessment?

- Is the assessment on one specific element of electrical safety?
  - E.G. Lock-out-tag-out
- If entire assessment is on only one topic it is described as being "*equally substitutable*" (Shrock & Coscarelli, 2007) so we can say with relative certainty that there is a gap in the knowledge of lock-out-tag-out in electrical safety
- If more than one area, we can only hypothesize that there is some sort of knowledge gap concerning electrical safety
- Evaluating by topic will present a more valid picture

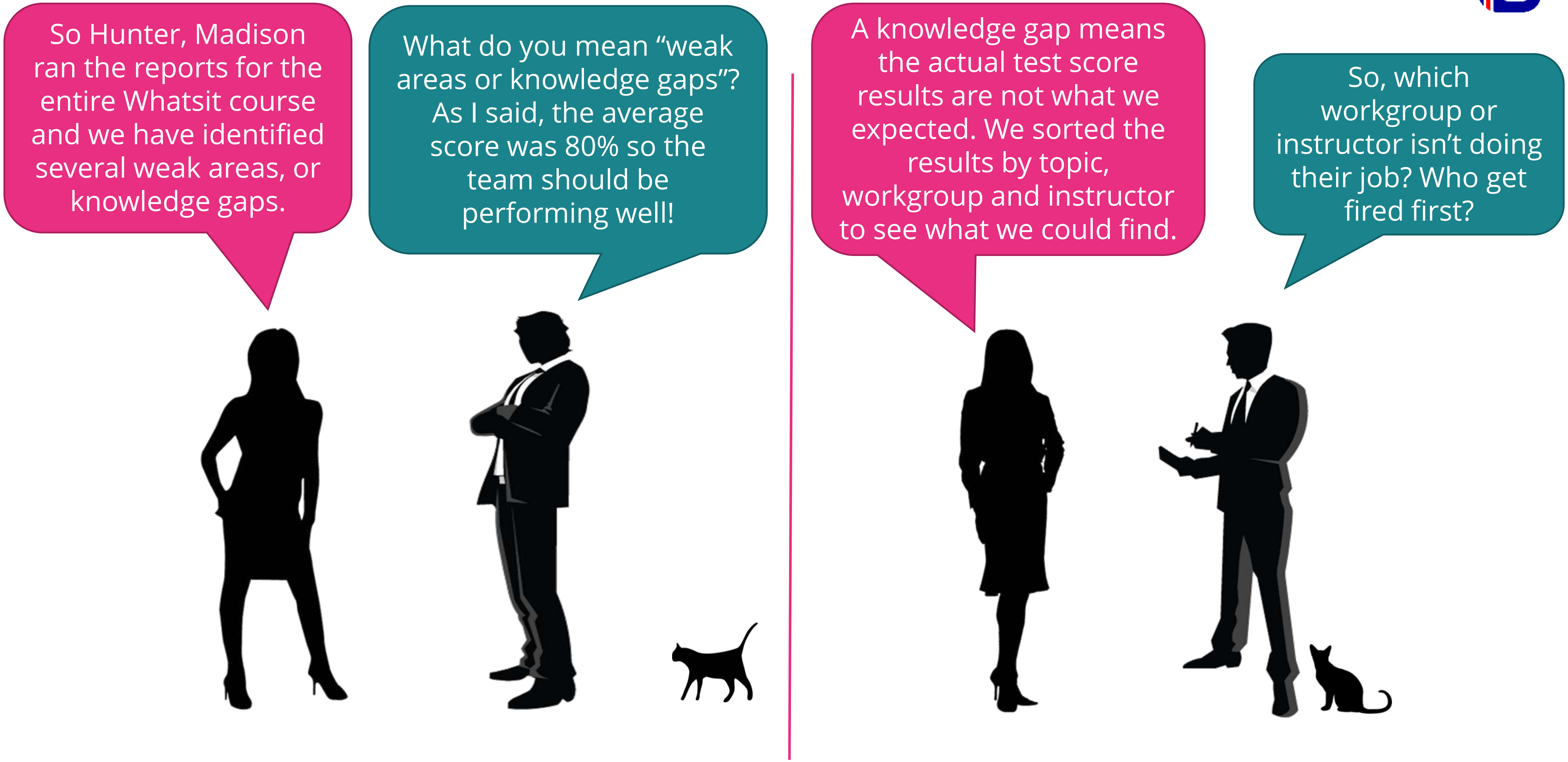
Actual Knowledge Gap Example

Average Scores by Topic on Electrical Safety Assessment

Topic	Expected Minimum Score	Average Score Attained	Probable Knowledge Gap
Lock-out-tag-out	100%	65%	35%
Grounding	80%	75%	5%
Insulation	70%	77%	**-7%**

Observed Results

- Assessment covered more than one topic
- Overall average score was 72.3%
  - Most participants "passed" using 60% standard
- Average scores by topic show very different results
  - Weak in lock-out-tag-out



So Hunter, Madison ran the reports for the entire Whatsit course and we have identified several weak areas, or knowledge gaps.

What do you mean “weak areas or knowledge gaps”? As I said, the average score was 80% so the team should be performing well!

A knowledge gap means the actual test score results are not what we expected. We sorted the results by topic, workgroup and instructor to see what we could find.

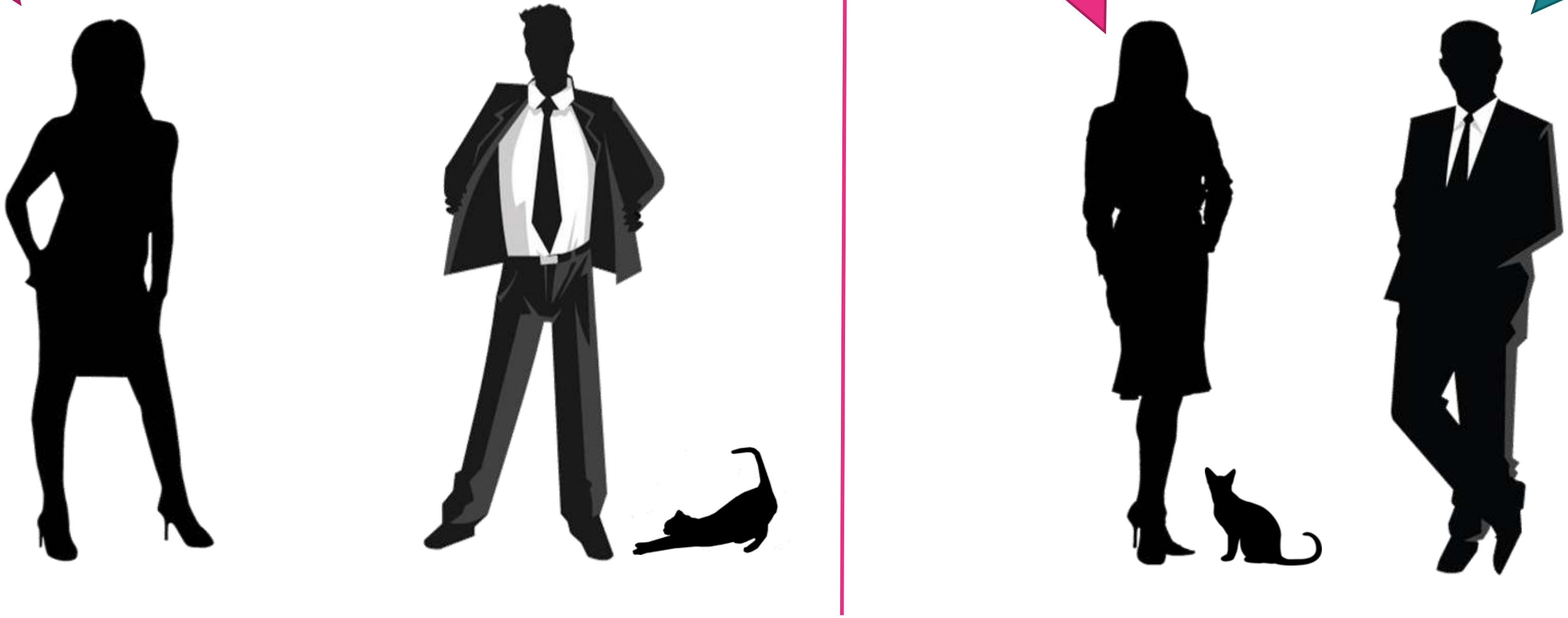
So, which workgroup or instructor isn’t doing their job? Who get fired first?

Nobody has to be fired Hunter, we just need to find the root cause of the poor scores in two topics; Whatsit troubleshooting and Whatsit error trapping. There may be a problem with the curriculum.

So what you're telling me Jess, is that it may be our fault because we are teaching the wrong things?

That is a strong possibility Hunter. The average scores on the other eight topics were pretty good. I think we should look at the curriculum to see what is actually being taught and tested.

Wow! Maybe it's my fault! I approved the curriculum. I guess I'll have to take a closer look to make sure we are teaching and testing the right things.



Plan Ahead

- It is important to plan ahead whether you want to be able look for a general knowledge gap or whether you want to break down the areas in which knowledge gaps occur
- When setting up your initial test item database think about what you want to know about the results
- Set up database by subject/topic
- Think about reporting by instructors or locations

question mark

Compass Consultants, LLC

Fair Test Design

Selecting Items by Topic and Difficulty

Quick Poll ☑

How does your organization select test-items for each assessment?

A. We use fixed-form exams – everyone gets same questions in same order

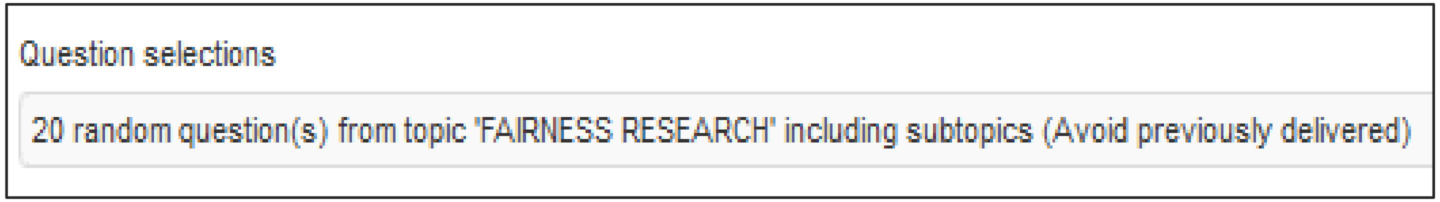
B. Items are selected randomly from all topics in the database

C. Random selection from all topics but specific number from each topic

D. Items are selected using stratified randomization by topic and difficulty to ensure fairness

Randomized Item Selection

- Experiments by Jim Parry:
  - Test-items selected at random from entire item database

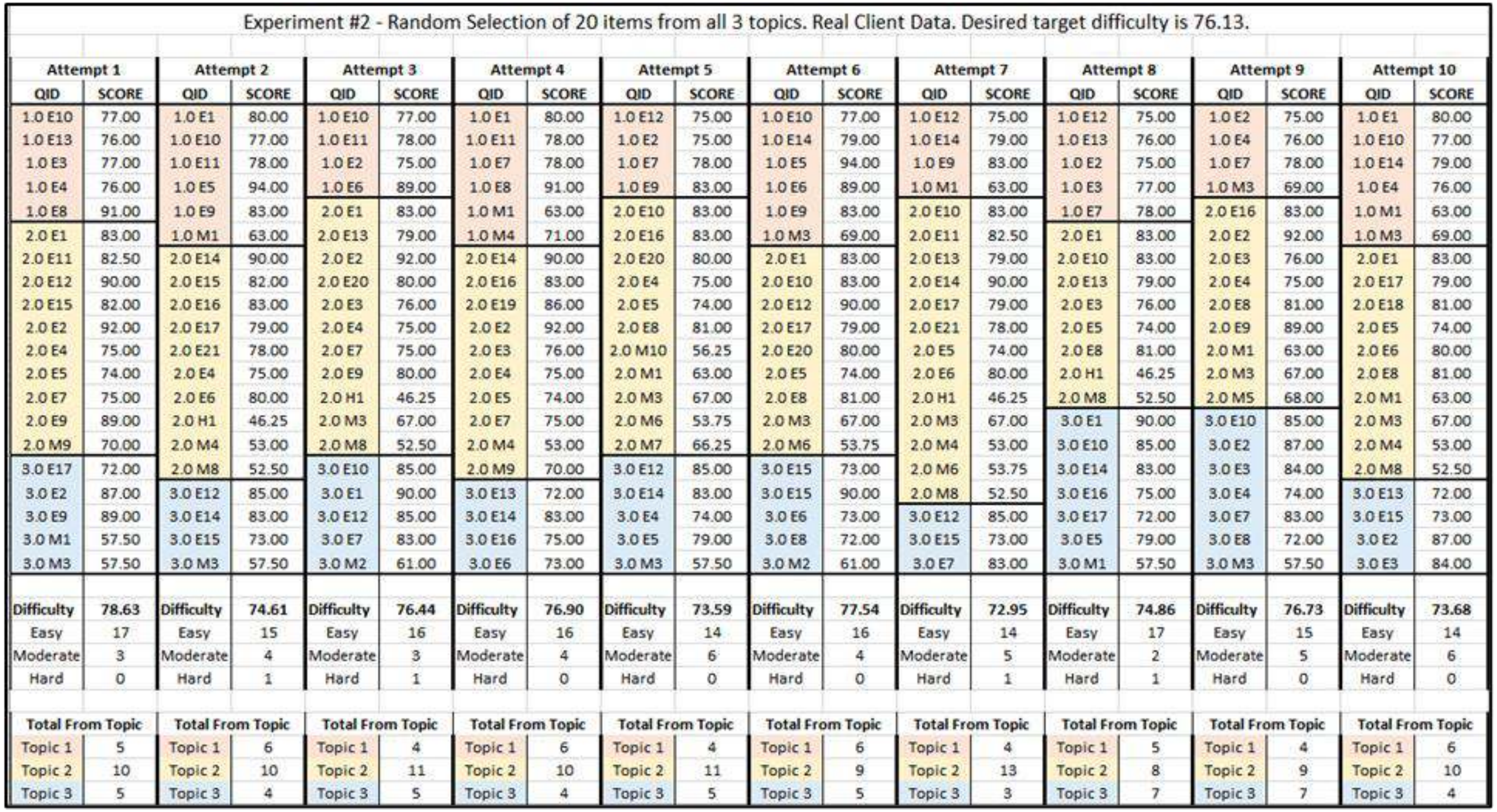


Question selections

20 random question(s) from topic 'FAIRNESS RESEARCH' including subtopics (Avoid previously delivered)

  - Produced unpredictable results in topic coverage although average difficulty was acceptable
    - Number of hard, moderate, and easy items varied significantly

Unpredictable Random Results



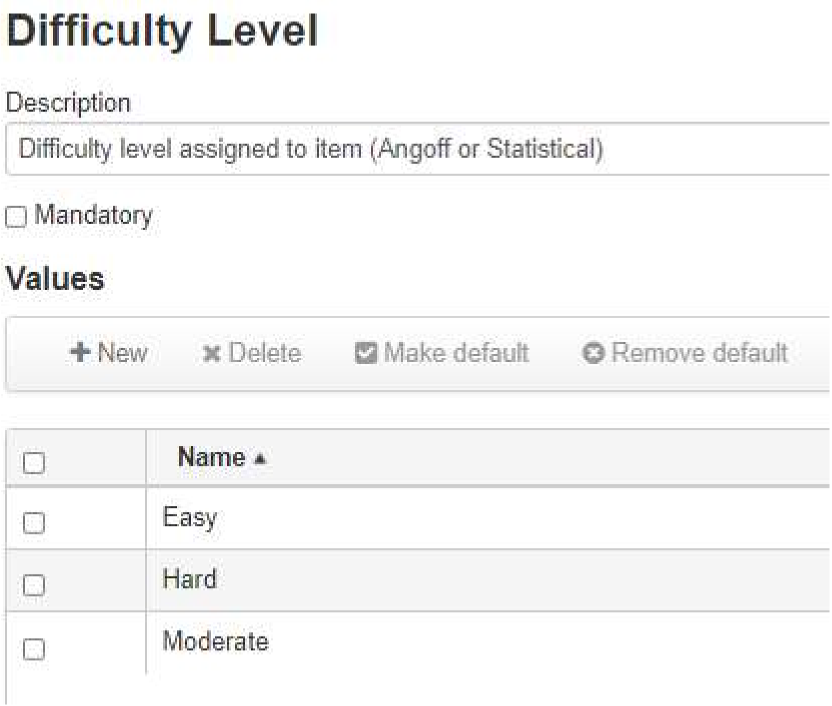
Experiment #2 - Random Selection of 20 items from all 3 topics. Real Client Data. Desired target difficulty is 76.13.

Attempt 1		Attempt 2		Attempt 3		Attempt 4		Attempt 5		Attempt 6		Attempt 7		Attempt 8		Attempt 9		Attempt 10	
QID	SCORE	QID	SCORE	QID	SCORE	QID	SCORE	QID	SCORE	QID	SCORE	QID	SCORE	QID	SCORE	QID	SCORE	QID	SCORE
1.0 E10	77.00	1.0 E1	80.00	1.0 E10	77.00	1.0 E1	80.00	1.0 E12	75.00	1.0 E10	77.00	1.0 E12	75.00	1.0 E12	75.00	1.0 E2	75.00	1.0 E1	80.00
1.0 E13	76.00	1.0 E10	77.00	1.0 E11	78.00	1.0 E11	78.00	1.0 E2	75.00	1.0 E14	79.00	1.0 E14	79.00	1.0 E13	76.00	1.0 E4	76.00	1.0 E10	77.00
1.0 E3	77.00	1.0 E11	78.00	1.0 E2	75.00	1.0 E7	78.00	1.0 E7	78.00	1.0 E5	94.00	1.0 E9	83.00	1.0 E2	75.00	1.0 E7	78.00	1.0 E14	79.00
1.0 E4	76.00	1.0 E5	94.00	1.0 E6	89.00	1.0 E8	91.00	1.0 E9	83.00	1.0 E6	89.00	1.0 M1	63.00	1.0 E3	77.00	1.0 M3	69.00	1.0 E4	76.00
1.0 E8	91.00	1.0 E9	83.00	2.0 E1	83.00	1.0 M1	63.00	2.0 E10	83.00	1.0 E9	83.00	2.0 E10	83.00	1.0 E7	78.00	2.0 E16	83.00	1.0 M1	63.00
2.0 E1	83.00	1.0 M1	63.00	2.0 E13	79.00	1.0 M4	71.00	2.0 E16	83.00	1.0 M3	69.00	2.0 E11	82.50	2.0 E1	83.00	2.0 E2	92.00	1.0 M3	69.00
2.0 E11	82.50	2.0 E14	90.00	2.0 E2	92.00	2.0 E14	90.00	2.0 E20	80.00	2.0 E1	83.00	2.0 E13	79.00	2.0 E10	83.00	2.0 E3	76.00	2.0 E1	83.00
2.0 E12	90.00	2.0 E15	82.00	2.0 E20	80.00	2.0 E16	83.00	2.0 E4	75.00	2.0 E10	83.00	2.0 E14	90.00	2.0 E13	79.00	2.0 E4	75.00	2.0 E17	79.00
2.0 E15	82.00	2.0 E16	83.00	2.0 E3	76.00	2.0 E19	86.00	2.0 E5	74.00	2.0 E12	90.00	2.0 E17	79.00	2.0 E3	76.00	2.0 E8	81.00	2.0 E18	81.00
2.0 E2	92.00	2.0 E17	79.00	2.0 E4	75.00	2.0 E2	92.00	2.0 E8	81.00	2.0 E17	79.00	2.0 E21	78.00	2.0 E5	74.00	2.0 E9	89.00	2.0 E5	74.00
2.0 E4	75.00	2.0 E21	78.00	2.0 E7	75.00	2.0 E3	76.00	2.0 M10	56.25	2.0 E20	80.00	2.0 E5	74.00	2.0 E8	81.00	2.0 M1	63.00	2.0 E6	80.00
2.0 E5	74.00	2.0 E4	75.00	2.0 E9	80.00	2.0 E4	75.00	2.0 M1	63.00	2.0 E5	74.00	2.0 E6	80.00	2.0 H1	46.25	2.0 M3	67.00	2.0 E8	81.00
2.0 E7	75.00	2.0 E6	80.00	2.0 H1	46.25	2.0 E5	74.00	2.0 M3	67.00	2.0 E8	81.00	2.0 H1	46.25	2.0 M8	52.50	2.0 M5	68.00	2.0 M1	63.00
2.0 E9	89.00	2.0 H1	46.25	2.0 M3	67.00	2.0 E7	75.00	2.0 M6	53.75	2.0 M3	67.00	2.0 M3	67.00	3.0 E1	90.00	3.0 E10	85.00	2.0 M3	67.00
2.0 M9	70.00	2.0 M4	53.00	2.0 M8	52.50	2.0 M4	53.00	2.0 M7	66.25	2.0 M6	53.75	2.0 M4	53.00	3.0 E10	85.00	3.0 E2	87.00	2.0 M4	53.00
3.0 E17	72.00	2.0 M8	52.50	3.0 E10	85.00	2.0 M9	70.00	3.0 E12	85.00	3.0 E15	73.00	2.0 M6	53.75	3.0 E14	83.00	3.0 E3	84.00	2.0 M8	52.50
3.0 E2	87.00	3.0 E12	85.00	3.0 E1	90.00	3.0 E13	72.00	3.0 E14	83.00	3.0 E15	90.00	2.0 M8	52.50	3.0 E16	75.00	3.0 E4	74.00	3.0 E13	72.00
3.0 E9	89.00	3.0 E14	83.00	3.0 E12	85.00	3.0 E14	83.00	3.0 E4	74.00	3.0 E6	73.00	3.0 E12	85.00	3.0 E17	72.00	3.0 E7	83.00	3.0 E15	73.00
3.0 M1	57.50	3.0 E15	73.00	3.0 E7	83.00	3.0 E16	75.00	3.0 E5	79.00	3.0 E8	72.00	3.0 E15	73.00	3.0 E5	79.00	3.0 E8	72.00	3.0 E2	87.00
3.0 M3	57.50	3.0 M3	57.50	3.0 M2	61.00	3.0 E6	73.00	3.0 M3	57.50	3.0 M2	61.00	3.0 E7	83.00	3.0 M1	57.50	3.0 M3	57.50	3.0 E3	84.00
Difficulty	78.63	Difficulty	74.61	Difficulty	76.44	Difficulty	76.90	Difficulty	73.59	Difficulty	77.54	Difficulty	72.95	Difficulty	74.86	Difficulty	76.73	Difficulty	73.68
Easy	17	Easy	15	Easy	16	Easy	16	Easy	14	Easy	16	Easy	14	Easy	17	Easy	15	Easy	14
Moderate	3	Moderate	4	Moderate	3	Moderate	4	Moderate	6	Moderate	4	Moderate	5	Moderate	2	Moderate	5	Moderate	6
Hard	0	Hard	1	Hard	1	Hard	0	Hard	0	Hard	0	Hard	1	Hard	1	Hard	0	Hard	0
Total From Topic		Total From Topic		Total From Topic		Total From Topic		Total From Topic		Total From Topic		Total From Topic		Total From Topic		Total From Topic		Total From Topic	
Topic 1	5	Topic 1	6	Topic 1	4	Topic 1	6	Topic 1	4	Topic 1	6	Topic 1	4	Topic 1	5	Topic 1	4	Topic 1	6
Topic 2	10	Topic 2	10	Topic 2	11	Topic 2	10	Topic 2	11	Topic 2	9	Topic 2	13	Topic 2	8	Topic 2	9	Topic 2	10
Topic 3	5	Topic 3	4	Topic 3	5	Topic 3	4	Topic 3	5	Topic 3	5	Topic 3	3	Topic 3	7	Topic 3	7	Topic 3	4

Set Up Database for Stratified Random Selection

- Populate test-items in database by objective, topic and difficulty
  - ○ Repository Name
    - Objective 1.0
      - Topic 1.1
        - Sub-Topic 1.1.1
          - 1.1.1 HARD
            - Test-Item 1.1.1/1
            - Test-Item 1.1.1/2
          - 1.1.1 MODERATE
            - Test-Item 1.1.1/3
            - Test-Item 1.1.1/4
          - 1.1.1 EASY
            - Test-Item 1.1.1/5
            - Test-Item 1.1.1/6

- Alternative – Use Metatags to identify difficulty of item



Difficulty Level

Description

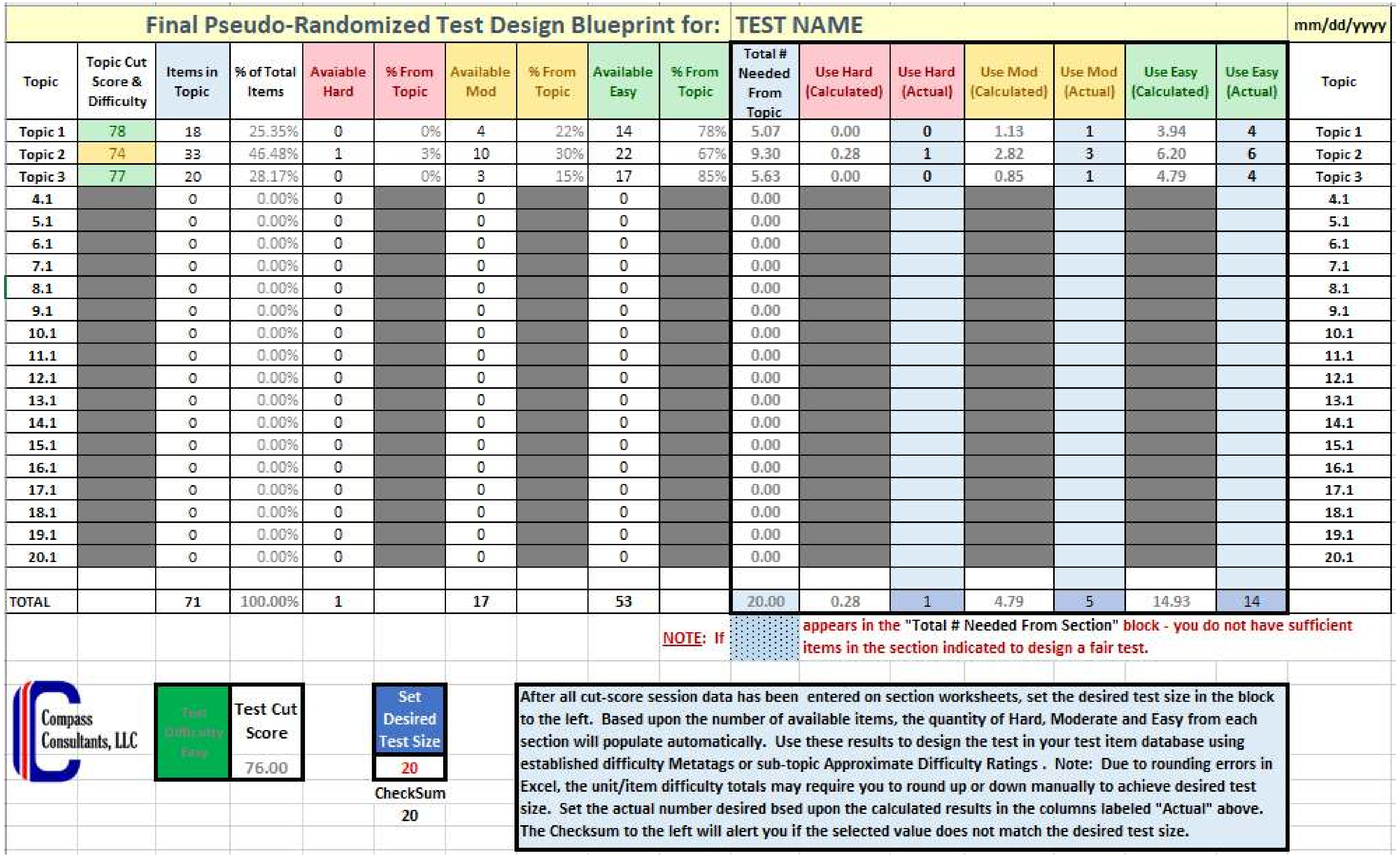
Difficulty level assigned to item (Angoff or Statistical)

□ Mandatory

Values

□	Name ▲
□	Easy
□	Hard
□	Moderate

Determine Stratification



Final Pseudo-Randomized Test Design Blueprint for: TEST NAME mm/dd/yyyy

Topic	Topic Cut Score & Difficulty	Items in Topic	% of Total Items	Avaiable Hard	% From Topic	Available Mod	% From Topic	Available Easy	% From Topic	Total # Needed From Topic	Use Hard (Calculated)	Use Hard (Actual)	Use Mod (Calculated)	Use Mod (Actual)	Use Easy (Calculated)	Use Easy (Actual)	Topic
Topic 1	78	18	25.35%	0	0%	4	22%	14	78%	5.07	0.00	0	1.13	1	3.94	4	Topic 1
Topic 2	74	33	46.48%	1	3%	10	30%	22	67%	9.30	0.28	1	2.82	3	6.20	6	Topic 2
Topic 3	77	20	28.17%	0	0%	3	15%	17	85%	5.63	0.00	0	0.85	1	4.79	4	Topic 3
4.1		0	0.00%	0		0		0		0.00							4.1
5.1		0	0.00%	0		0		0		0.00							5.1
6.1		0	0.00%	0		0		0		0.00							6.1
7.1		0	0.00%	0		0		0		0.00							7.1
8.1		0	0.00%	0		0		0		0.00							8.1
9.1		0	0.00%	0		0		0		0.00							9.1
10.1		0	0.00%	0		0		0		0.00							10.1
11.1		0	0.00%	0		0		0		0.00							11.1
12.1		0	0.00%	0		0		0		0.00							12.1
13.1		0	0.00%	0		0		0		0.00							13.1
14.1		0	0.00%	0		0		0		0.00							14.1
15.1		0	0.00%	0		0		0		0.00							15.1
16.1		0	0.00%	0		0		0		0.00							16.1
17.1		0	0.00%	0		0		0		0.00							17.1
18.1		0	0.00%	0		0		0		0.00							18.1
19.1		0	0.00%	0		0		0		0.00							19.1
20.1		0	0.00%	0		0		0		0.00							20.1
TOTAL		71	100.00%	1		17		53		20.00	0.28	1	4.79	5	14.93	14	

NOTE: If ▨ appears in the "Total # Needed From Section" block - you do not have sufficient items in the section indicated to design a fair test.

Compass Consultants, LLC

Test Difficulty Easy	Test Cut Score
	76.00

Set Desired Test Size

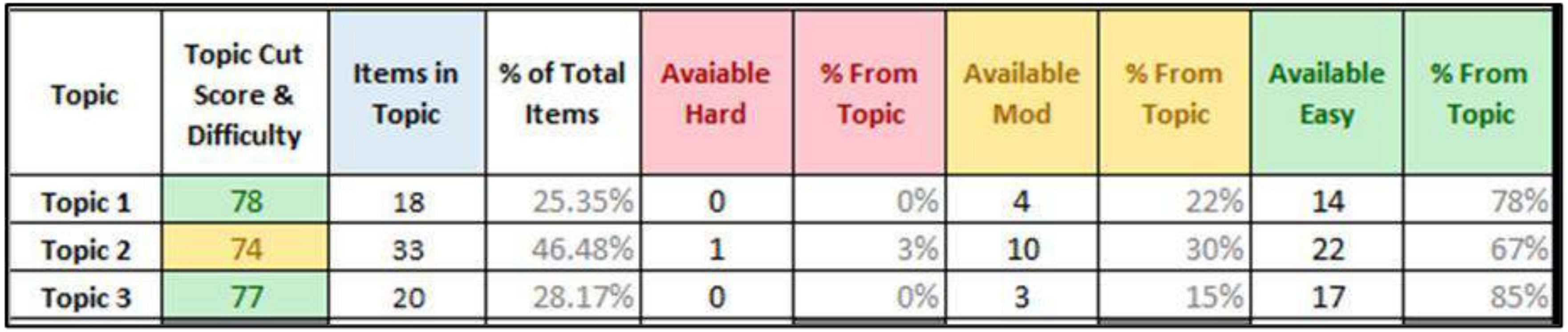
20

CheckSum

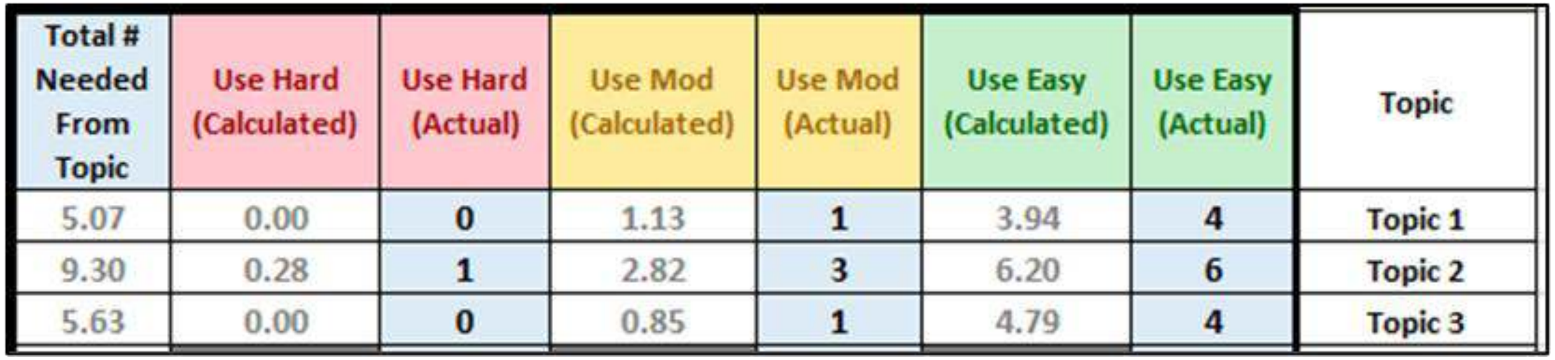
20

After all cut-score session data has been entered on section worksheets, set the desired test size in the block to the left. Based upon the number of available items, the quantity of Hard, Moderate and Easy from each section will populate automatically. Use these results to design the test in your test item database using established difficulty Metatags or sub-topic Approximate Difficulty Ratings . Note: Due to rounding errors in Excel, the unit/item difficulty totals may require you to round up or down manually to achieve desired test size. Set the actual number desired bsed upon the calculated results in the columns labeled "Actual" above. The Checksum to the left will alert you if the selected value does not match the desired test size.

Total test-items available by topic at each difficulty level.



Topic	Topic Cut Score & Difficulty	Items in Topic	% of Total Items	Avaiable Hard	% From Topic	Available Mod	% From Topic	Available Easy	% From Topic
Topic 1	78	18	25.35%	0	0%	4	22%	14	78%
Topic 2	74	33	46.48%	1	3%	10	30%	22	67%
Topic 3	77	20	28.17%	0	0%	3	15%	17	85%

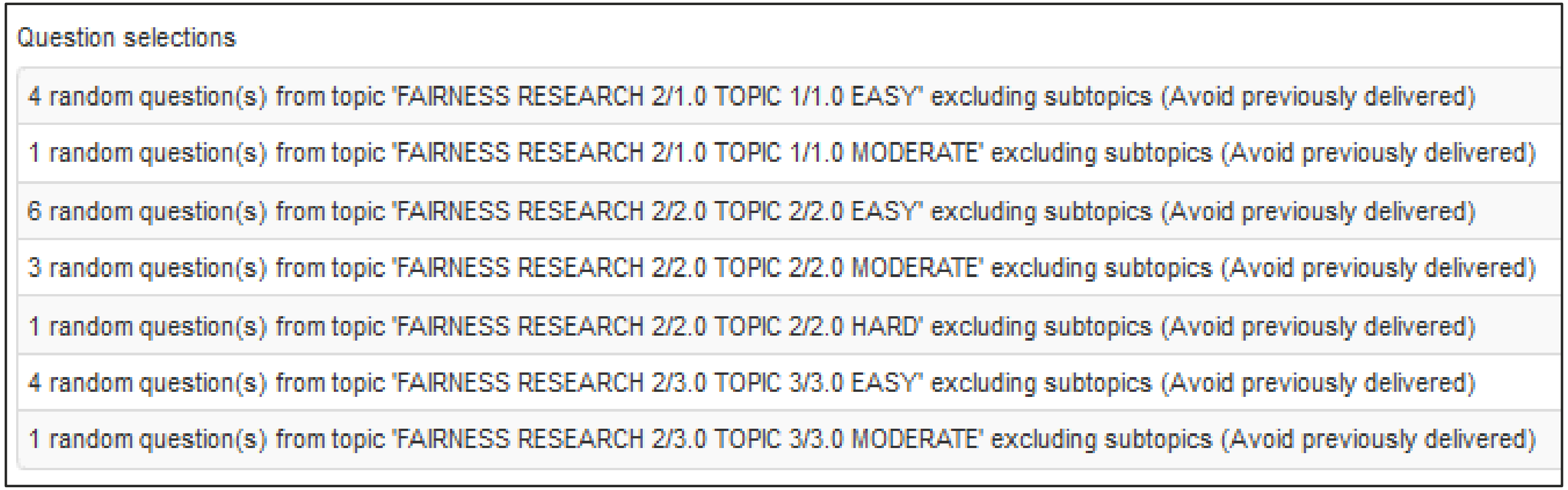


Total # Needed From Topic	Use Hard (Calculated)	Use Hard (Actual)	Use Mod (Calculated)	Use Mod (Actual)	Use Easy (Calculated)	Use Easy (Actual)	Topic
5.07	0.00	0	1.13	1	3.94	4	Topic 1
9.30	0.28	1	2.82	3	6.20	6	Topic 2
5.63	0.00	0	0.85	1	4.79	4	Topic 3

Recommended test design based on number of items available at each difficulty level to maintain difficulty and topic coverage.

Stratified Random Item Selection Criteria

- Test-items selected by both topic and difficulty



Question selections

4 random question(s) from topic 'FAIRNESS RESEARCH 2/1.0 TOPIC 1/1.0 EASY' excluding subtopics (Avoid previously delivered)

1 random question(s) from topic 'FAIRNESS RESEARCH 2/1.0 TOPIC 1/1.0 MODERATE' excluding subtopics (Avoid previously delivered)

6 random question(s) from topic 'FAIRNESS RESEARCH 2/2.0 TOPIC 2/2.0 EASY' excluding subtopics (Avoid previously delivered)

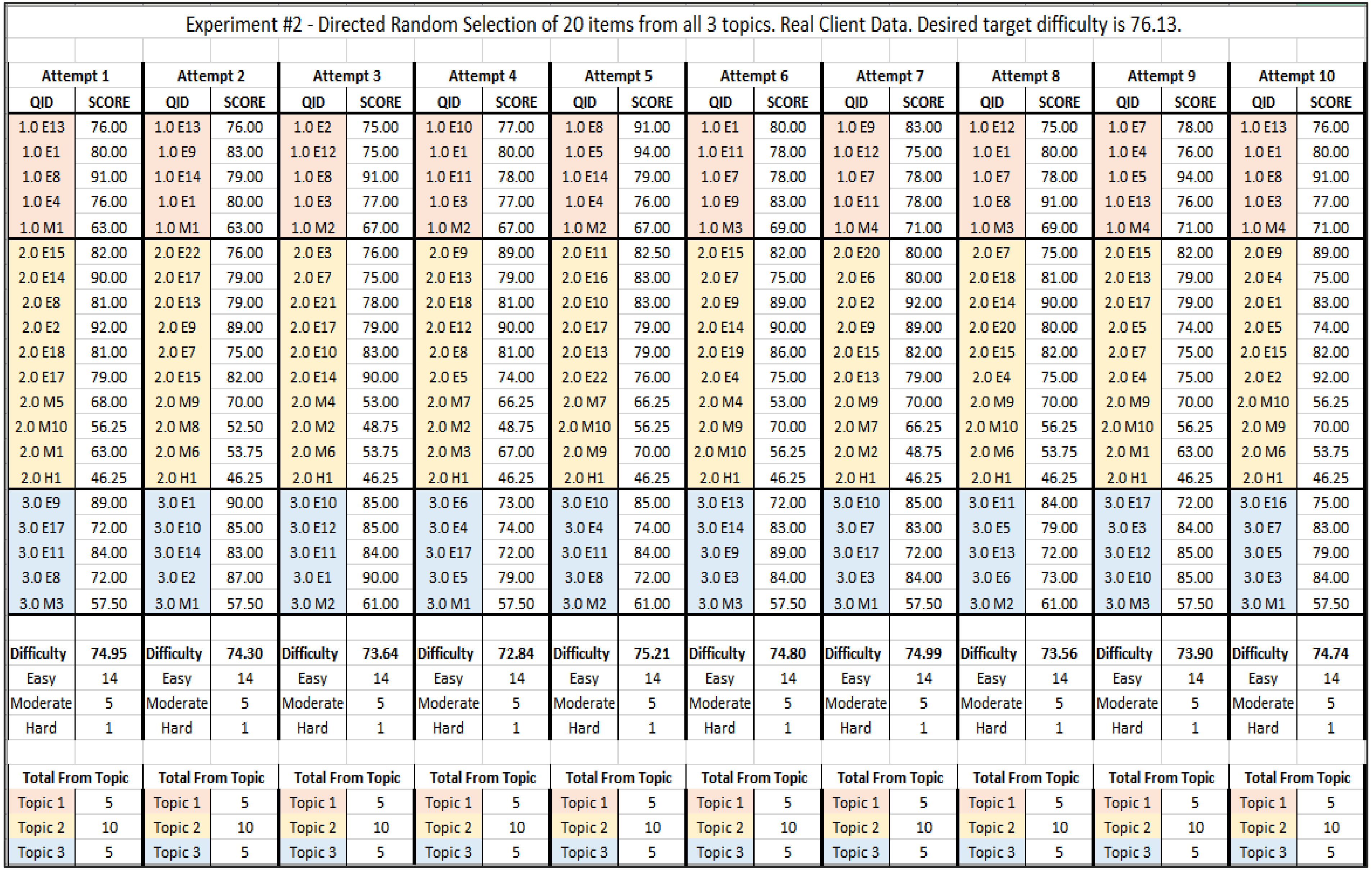
3 random question(s) from topic 'FAIRNESS RESEARCH 2/2.0 TOPIC 2/2.0 MODERATE' excluding subtopics (Avoid previously delivered)

1 random question(s) from topic 'FAIRNESS RESEARCH 2/2.0 TOPIC 2/2.0 HARD' excluding subtopics (Avoid previously delivered)

4 random question(s) from topic 'FAIRNESS RESEARCH 2/3.0 TOPIC 3/3.0 EASY' excluding subtopics (Avoid previously delivered)

1 random question(s) from topic 'FAIRNESS RESEARCH 2/3.0 TOPIC 3/3.0 MODERATE' excluding subtopics (Avoid previously delivered)

- Produces same topic coverage and acceptable test difficulty for each test
  - Number of hard, moderate, and easy items remains constant
  - Difficulty remains near the target within acceptable tolerance



Experiment #2 - Directed Random Selection of 20 items from all 3 topics. Real Client Data. Desired target difficulty is 76.13.

Attempt 1		Attempt 2		Attempt 3		Attempt 4		Attempt 5	
QID	SCORE	QID	SCORE	QID	SCORE	QID	SCORE	QID	SCORE
1.0 E13	76.00	1.0 E13	76.00	1.0 E2	75.00	1.0 E10	77.00	1.0 E8	91.00
1.0 E1	80.00	1.0 E9	83.00	1.0 E12	75.00	1.0 E1	80.00	1.0 E5	94.00
1.0 E8	91.00	1.0 E14	79.00	1.0 E8	91.00	1.0 E11	78.00	1.0 E14	79.00
1.0 E4	76.00	1.0 E1	80.00	1.0 E3	77.00	1.0 E3	77.00	1.0 E4	76.00
1.0 M1	63.00	1.0 M1	63.00	1.0 M2	67.00	1.0 M2	67.00	1.0 M2	67.00
2.0 E15	82.00	2.0 E22	76.00	2.0 E3	76.00	2.0 E9	89.00	2.0 E11	82.50
2.0 E14	90.00	2.0 E17	79.00	2.0 E7	75.00	2.0 E13	79.00	2.0 E16	83.00
2.0 E8	81.00	2.0 E13	79.00	2.0 E21	78.00	2.0 E18	81.00	2.0 E10	83.00
2.0 E2	92.00	2.0 E9	89.00	2.0 E17	79.00	2.0 E12	90.00	2.0 E17	79.00
2.0 E18	81.00	2.0 E7	75.00	2.0 E10	83.00	2.0 E8	81.00	2.0 E13	79.00
2.0 E17	79.00	2.0 E15	82.00	2.0 E14	90.00	2.0 E5	74.00	2.0 E22	76.00
2.0 M5	68.00	2.0 M9	70.00	2.0 M4	53.00	2.0 M7	66.25	2.0 M7	66.25
2.0 M10	56.25	2.0 M8	52.50	2.0 M2	48.75	2.0 M2	48.75	2.0 M10	56.25
2.0 M1	63.00	2.0 M6	53.75	2.0 M6	53.75	2.0 M3	67.00	2.0 M9	70.00
2.0 H1	46.25	2.0 H1	46.25	2.0 H1	46.25	2.0 H1	46.25	2.0 H1	46.25
3.0 E9	89.00	3.0 E1	90.00	3.0 E10	85.00	3.0 E6	73.00	3.0 E10	85.00
3.0 E17	72.00	3.0 E10	85.00	3.0 E12	85.00	3.0 E4	74.00	3.0 E4	74.00
3.0 E11	84.00	3.0 E14	83.00	3.0 E11	84.00	3.0 E17	72.00	3.0 E11	84.00
3.0 E8	72.00	3.0 E2	87.00	3.0 E1	90.00	3.0 E5	79.00	3.0 E8	72.00
3.0 M3	57.50	3.0 M1	57.50	3.0 M2	61.00	3.0 M1	57.50	3.0 M2	61.00
Difficulty	74.95	Difficulty	74.30	Difficulty	73.64	Difficulty	72.84	Difficulty	75.21
Easy	14	Easy	14	Easy	14	Easy	14	Easy	14
Moderate	5	Moderate	5	Moderate	5	Moderate	5	Moderate	5
Hard	1	Hard	1	Hard	1	Hard	1	Hard	1
Total From Topic		Total From Topic		Total From Topic		Total From Topic		Total From Topic	
Topic 1	5	Topic 1	5	Topic 1	5	Topic 1	5	Topic 1	5
Topic 2	10	Topic 2	10	Topic 2	10	Topic 2	10	Topic 2	10
Topic 3	5	Topic 3	5	Topic 3	5	Topic 3	5	Topic 3	5

Attempt 6		Attempt 7		Attempt 8		Attempt 9		Attempt 10	
QID	SCORE	QID	SCORE	QID	SCORE	QID	SCORE	QID	SCORE
1.0 E1	80.00	1.0 E9	83.00	1.0 E12	75.00	1.0 E7	78.00	1.0 E13	76.00
1.0 E11	78.00	1.0 E12	75.00	1.0 E1	80.00	1.0 E4	76.00	1.0 E1	80.00
1.0 E7	78.00	1.0 E7	78.00	1.0 E7	78.00	1.0 E5	94.00	1.0 E8	91.00
1.0 E9	83.00	1.0 E11	78.00	1.0 E8	91.00	1.0 E13	76.00	1.0 E3	77.00
1.0 M3	69.00	1.0 M4	71.00	1.0 M3	69.00	1.0 M4	71.00	1.0 M4	71.00
2.0 E15	82.00	2.0 E20	80.00	2.0 E7	75.00	2.0 E15	82.00	2.0 E9	89.00
2.0 E7	75.00	2.0 E6	80.00	2.0 E18	81.00	2.0 E13	79.00	2.0 E4	75.00
2.0 E9	89.00	2.0 E2	92.00	2.0 E14	90.00	2.0 E17	79.00	2.0 E1	83.00
2.0 E14	90.00	2.0 E9	89.00	2.0 E20	80.00	2.0 E5	74.00	2.0 E5	74.00
2.0 E19	86.00	2.0 E15	82.00	2.0 E15	82.00	2.0 E7	75.00	2.0 E15	82.00
2.0 E4	75.00	2.0 E13	79.00	2.0 E4	75.00	2.0 E4	75.00	2.0 E2	92.00
2.0 M4	53.00	2.0 M9	70.00	2.0 M9	70.00	2.0 M9	70.00	2.0 M10	56.25
2.0 M9	70.00	2.0 M7	66.25	2.0 M10	56.25	2.0 M10	56.25	2.0 M9	70.00
2.0 M10	56.25	2.0 M2	48.75	2.0 M6	53.75	2.0 M1	63.00	2.0 M6	53.75
2.0 H1	46.25	2.0 H1	46.25	2.0 H1	46.25	2.0 H1	46.25	2.0 H1	46.25
3.0 E13	72.00	3.0 E10	85.00	3.0 E11	84.00	3.0 E17	72.00	3.0 E16	75.00
3.0 E14	83.00	3.0 E7	83.00	3.0 E5	79.00	3.0 E3	84.00	3.0 E7	83.00
3.0 E9	89.00	3.0 E17	72.00	3.0 E13	72.00	3.0 E12	85.00	3.0 E5	79.00
3.0 E3	84.00	3.0 E3	84.00	3.0 E6	73.00	3.0 E10	85.00	3.0 E3	84.00
3.0 M3	57.50	3.0 M1	57.50	3.0 M2	61.00	3.0 M3	57.50	3.0 M1	57.50
Difficulty	74.80	Difficulty	74.99	Difficulty	73.56	Difficulty	73.90	Difficulty	74.74
Easy	14	Easy	14	Easy	14	Easy	14	Easy	14
Moderate	5	Moderate	5	Moderate	5	Moderate	5	Moderate	5
Hard	1	Hard	1	Hard	1	Hard	1	Hard	1
Total From Topic		Total From Topic		Total From Topic		Total From Topic		Total From Topic	
Topic 1	5	Topic 1	5	Topic 1	5	Topic 1	5	Topic 1	5
Topic 2	10	Topic 2	10	Topic 2	10	Topic 2	10	Topic 2	10
Topic 3	5	Topic 3	5	Topic 3	5	Topic 3	5	Topic 3	5

question mark

Compass Consultants, LLC

Scoring the Assessment

All at Once or by Topic?

Quick Poll ☑

How does your organization score assessments?

A. Our assessments contain only one topic so they are scored as a whole

B. We score the assessment as a whole even if it contains several topics

C. Topics are scored separately and student must pass all topics to pass

D. Topics are scored individually but only final average score matters

Your Call!

- Remember the knowledge gap identification!
- If all one topic – score entire assessment
- If several topics – possibly score individually
  - Topic scores may be averaged to obtain final score
  - Student may be required to pass all or some topics to achieve a passing score
    - Certain topics may have different passing/cut-score levels
- Topic level feedback is important
  - Test-takers
  - Instructors
  - Administrators
  - Location

Good Topic Structure + Intelligent Reporting = Valid Results

question mark

Compass Consultants, LLC

Questions?

White papers, infographics, reports, eBooks and more!

VIEW NOW:

White Paper: Assessment Results You Can Trust: https://www.questionmark.com/download-assessment-results-you-can-trust/

Book Review: *Criterion-Referenced Test Development*: https://www.questionmark.com/wp-content/uploads/2020/11/Book-Review-Criterion-Referenced-Test-1.pdf

Upcoming webinars

Item Writing: Tips & Techniques for Writing Good Questions

- September 14, 2021 - 11:00 am to 12:00 pm (EDT)

This webinar provides helpful guidance and tips for people who are are new to item writing, or those who are looking for ideas on coaching subject matter experts (SMEs) on item writing techniques.

Click to Register

Introduction to Questionmark's Assessment Platform

- September 16, 2021 - 10:00 am to 11:00 am (EDT)

Learn the basics of authoring, delivering and reporting on surveys, quizzes, tests and exams. This introductory webinar explains and demonstrates key Questionmark features and functions.

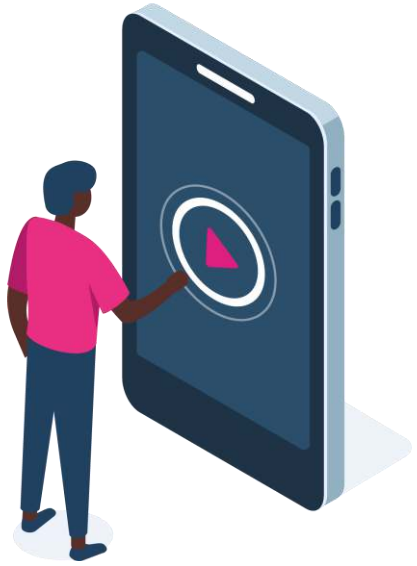
Click to Register

Making Scores Meaningful: The Role of Standards

- September 23, 2021 - 11:00 am to 12:00 pm (EDT)

This session will explore some key concepts in understanding assessment practices. Using examples from the driving test, to educational qualifications, to job interviews, we can see the real-world applicability of these concepts, and see their relevance to one's own assessment decision making:

Click to Register



Thank you for attending!

Reach out to Questionmark at sales@questionmark.com
or request a demo at https://www.questionmark.com/request-demo

If you would like to reach out to Jim Parry – james.parry@gocompassconsultants.com
www.gocompassconsultants.com