Identifying The Knowledge Gap

James R. Parry, M.Ed. CPT
Owner/Chief Executive Manager
Compass Consultants, LLC
References


Questionmark (2017), *What does the score list report do?* Retrieved from Questionmark.com: [https://www.questionmark.com/content/what-does-score-list-report-do](https://www.questionmark.com/content/what-does-score-list-report-do)

Acknowledgments

• Photos and some illustrations used fall within the creative commons license
• Business men/women figures courtesy of E-Learning Heroes
Let’s listen in...
Stacy, we seem to have a knowledge problem among the team working on the Snapperdoodle development project!

What makes you say that Steve?
Well, they just don’t seem to understand! They should know what we expect of them!

Has the team lead made them aware of her expectations?
Of course! They all went to the same training sessions and the average final test score at the end was in the 80s. That tells me that they are aware of the expectations!

Steve, 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 in the 80s so that tells me they know what all they have to know!

Remember, the overall average just gives a snapshot of the whole test, Steve, but it doesn’t show which specific areas or topics are deficient.
Yeah, I do seem to remember hearing something like that. Is there any way you can see if there are specific weak areas?

Sure thing Steve! I will run some analytics in the Questionmark application and see what I can come up with. Have a great day!
Hi Eric. I just had a discussion with Stacy. She said she would try to identify what’s wrong with the team by running something called analytics in some program called Questionmark.
Hi Julie. Could you run the analytics on the quizzes and the final tests for the Snapperdoodle project classes. Please be sure to report by topic. We need to see if there are specific weaknesses in the training.

No problem Stacy! I’ll have the results by the time you get back from lunch. You know as well as I do that the analytics that Questionmark On-Demand produces will tell us the whole story!
So...
Knowledge

“Information, understanding, or skill that you get from experience or education”

“Awareness of something: the state of being aware of “something””

Knowledge, 2016
Gap

“An incomplete or deficient area <a gap in her knowledge>”

Merriam-Webster, 2016
“A disparity in levels of (especially technological) knowledge”
Is There a 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 or mission, how can we identify where knowledge gaps exist?
Assumption

• For the purpose of this discussion, we will consider the definition of knowledge gap within the context of criterion-referenced testing of knowledge
  • This is as opposed to testing of performance or to norm referenced testing
“...it is important to remember that the only legitimate score to report is “master” or “non-master”.

“CRTs are designed to measure specific competencies and to sort test takers into two classifications, that is, masters and non-masters. These tests are not designed to separate test-takers reliably into graduations of master or levels of non-master. In other words, these tests (unlike norm-referenced tests) are not designed to separate and compare test takers to one another.”

Shrock & Coscarelli (2007)
Expected Level of Knowledge

• Remember the formative years of middle & high school as well as most colleges
• “Expected” level of knowledge was defined by letter grades associated with a range of numbers grades

<table>
<thead>
<tr>
<th>Letter Grade</th>
<th>Score Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>90 – 100%</td>
</tr>
<tr>
<td>B</td>
<td>80 – 89%</td>
</tr>
<tr>
<td>C</td>
<td>70 – 79%</td>
</tr>
<tr>
<td>D</td>
<td>60 – 69%</td>
</tr>
<tr>
<td>F</td>
<td>&lt;60%</td>
</tr>
</tbody>
</table>
What it Means

• If you do not achieve a 60% average, you have not met the minimum “expected” requirements
  • You fail!

• Grade generally represents a percentage of knowledge that a student can regurgitate from memory on some sort of written assessment
False Knowledge Gap

• Don’t make assumptions about the “gap” based on incomplete information

• If test covers only one topic/subject the gap may exist due to incorrect information
  • Objectives may not match expected performance
False Knowledge Gap

Scores on electrical safety assessment

<table>
<thead>
<tr>
<th>Number Attaining Score</th>
<th>Score Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>100%</td>
</tr>
<tr>
<td>5</td>
<td>90 – 99%</td>
</tr>
<tr>
<td>15</td>
<td>80 – 89%</td>
</tr>
<tr>
<td>20</td>
<td>70 – 79%</td>
</tr>
<tr>
<td>20</td>
<td>60 – 69%</td>
</tr>
<tr>
<td>35</td>
<td>&lt;60%</td>
</tr>
</tbody>
</table>

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
Alternative False Knowledge Gap

Scores on electrical safety assessment

<table>
<thead>
<tr>
<th>Number Attaining Score</th>
<th>Score Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>100%</td>
</tr>
<tr>
<td>5</td>
<td>90 – 99%</td>
</tr>
<tr>
<td>15</td>
<td>80 – 89%</td>
</tr>
<tr>
<td>20</td>
<td>70 – 79%</td>
</tr>
<tr>
<td>20</td>
<td>60 – 69%</td>
</tr>
<tr>
<td>35</td>
<td>&lt;60%</td>
</tr>
</tbody>
</table>

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%

Copyright © 2017 by Compass Consultants, LLC
What’s 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 knowledge 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
Actual Knowledge Gap

**Average scores by topic on electrical safety assessment**

<table>
<thead>
<tr>
<th>Topic</th>
<th>Expected Score</th>
<th>Average Score</th>
<th>Knowledge Gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lock-out-tag-out</td>
<td>100%</td>
<td>65%</td>
<td>35%</td>
</tr>
<tr>
<td>Grounding</td>
<td>80%</td>
<td>75%</td>
<td>5%</td>
</tr>
<tr>
<td>Insulation</td>
<td>92%</td>
<td>77%</td>
<td>15%</td>
</tr>
</tbody>
</table>

- 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
Expected Passing Score

• What does “expected passing score” actually mean and how is it determined?

• Let’s take for example the “expected passing score” in the antiquated universal scoring system
  • Why is 60% and above a passing score?
  • Is it okay for a student to only retain 60% of every bit of knowledge?
  • Why not 100%, 82%, 76%, or 45%, etc.?
  • Who decides this and how is it determined?
  • Is there a better way to set a passing or cut score that will more accurately reflect actual knowledge retention?
Setting an Achievable Cut Score

Don’t just pull a rabbit out of the hat!
Identified Knowledge Gap Needs Meaning

• Must use trustworthy method to determine “expected” or “cut” score
  • Must actually be achievable

• Three accepted methods for criterion-referenced tests
  • Informed judgement
  • Conjectural (Modified Angoff)
  • Contrasting groups

• One NOT so acceptable but used most often
  • Arbitrary method
The Arbitrary Method

• In this familiar method someone (person, organization, etc.) decides that 60% or 80%, or 100%, etc. should be passing, based upon nothing other than supposition that this is what is “good enough”

• This method is not recommended because it may result in false classification of candidates
  • This false classification can be either false-positive or false-negative.
    • An example of a false-positive error would be an examinee who was not minimally competent, but who passed the test
    • In a false-negative situation an examinee who actually has the level of competence required, fails the test

• Using an arbitrary method to set a passing or cut score can produce erroneous results that may lead to false conclusions

(Professional Testing, Inc., 2016)
Arbitrary
Informed Judgement Method

• This is what is called a test-based approach
  • A panel of judges, who may be stakeholders in the course of instruction, review each test and its content
  • Each judge suggests a percentage of items that he or she believes a minimally competent candidate ought to answer correctly
  • The decision maker then reviews the judges recommendations, along with additional factors and sets the cut or passing score
  • This score may be difficult to rationally defend unless it is used in combination with other methods, particularly the contrasting groups method

(Professional Testing, Inc., 2016)
Informed Judgement

<table>
<thead>
<tr>
<th>Subject</th>
<th>Mat</th>
<th>Ger</th>
<th>Bio</th>
<th>Che</th>
<th>Phy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mat</td>
<td></td>
<td>.58*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ger</td>
<td>.58*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bio</td>
<td>.58*</td>
<td>.59*</td>
<td>.67*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Che</td>
<td>.65*</td>
<td>.57*</td>
<td>.67*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phy</td>
<td>.64*</td>
<td>.56*</td>
<td>.65*</td>
<td>.68*</td>
<td></td>
</tr>
</tbody>
</table>
Conjectural (Modified Angoff) Method

• The Modified Angoff method is the most commonly used of the conjectural methods when setting cut scores for high-stakes examinations
• This is an item-based approach to standard setting
• A panel of judges is assembled and asked to review each test item, one at a time
• As they review the items they estimate the percentage of a hypothetical group (usually 100) of minimally competent performers or examinees who would likely respond correctly to the given item
• The item probabilities or percentages are then averaged for each judge and are averaged across the entire set of test items to produce a recommended cut or passing score
## Modified Angoff Method

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>7.1.1.1/1</td>
<td>Moderate</td>
<td>72.00</td>
<td>70</td>
<td>80</td>
<td>60</td>
<td>75</td>
<td>75</td>
<td>75</td>
<td>7.58</td>
<td>Moderate</td>
</tr>
<tr>
<td>7.1.1.1/2</td>
<td>Moderate</td>
<td>72.00</td>
<td>70</td>
<td>75</td>
<td>55</td>
<td>80</td>
<td>80</td>
<td>80</td>
<td>10.37</td>
<td></td>
</tr>
<tr>
<td>7.1.1.1/3</td>
<td>Easy</td>
<td>79.00</td>
<td>85</td>
<td>80</td>
<td>70</td>
<td>90</td>
<td>70</td>
<td>70</td>
<td>8.94</td>
<td></td>
</tr>
<tr>
<td>7.1.1.1/4</td>
<td>Moderate</td>
<td>60.00</td>
<td>70</td>
<td>60</td>
<td>40</td>
<td>60</td>
<td>70</td>
<td>70</td>
<td>12.25</td>
<td></td>
</tr>
<tr>
<td>7.1.1.2/1</td>
<td>Hard</td>
<td>49.00</td>
<td>55</td>
<td>50</td>
<td>45</td>
<td>40</td>
<td>55</td>
<td>55</td>
<td>6.52</td>
<td></td>
</tr>
<tr>
<td>7.1.1.2/2</td>
<td>Moderate</td>
<td>53.00</td>
<td>60</td>
<td>50</td>
<td>45</td>
<td>50</td>
<td>60</td>
<td>60</td>
<td>6.71</td>
<td></td>
</tr>
<tr>
<td>7.1.1.2/3</td>
<td>Moderate</td>
<td>52.00</td>
<td>60</td>
<td>50</td>
<td>35</td>
<td>55</td>
<td>55</td>
<td>55</td>
<td>10.37</td>
<td></td>
</tr>
<tr>
<td>7.1.1.2/4</td>
<td>Moderate</td>
<td>63.00</td>
<td>75</td>
<td>60</td>
<td>45</td>
<td>60</td>
<td>75</td>
<td>75</td>
<td>12.55</td>
<td></td>
</tr>
<tr>
<td>7.1.1.2/5</td>
<td>Moderate</td>
<td>73.00</td>
<td>85</td>
<td>70</td>
<td>60</td>
<td>75</td>
<td>75</td>
<td>75</td>
<td>9.08</td>
<td></td>
</tr>
<tr>
<td>7.1.1.2/6</td>
<td>Moderate</td>
<td>74.00</td>
<td>85</td>
<td>75</td>
<td>70</td>
<td>80</td>
<td>80</td>
<td>80</td>
<td>9.62</td>
<td></td>
</tr>
<tr>
<td>7.1.1.2/7</td>
<td>Moderate</td>
<td>74.00</td>
<td>85</td>
<td>75</td>
<td>70</td>
<td>70</td>
<td>70</td>
<td>70</td>
<td>6.52</td>
<td></td>
</tr>
<tr>
<td>7.1.1.2/8</td>
<td>Moderate</td>
<td>68.00</td>
<td>75</td>
<td>75</td>
<td>70</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>7.58</td>
<td></td>
</tr>
<tr>
<td>7.1.1.2/9</td>
<td>Moderate</td>
<td>64.00</td>
<td>75</td>
<td>75</td>
<td>55</td>
<td>60</td>
<td>55</td>
<td>55</td>
<td>10.25</td>
<td></td>
</tr>
<tr>
<td>7.1.1.3/1</td>
<td>Moderate</td>
<td>61.00</td>
<td>70</td>
<td>60</td>
<td>55</td>
<td>50</td>
<td>70</td>
<td>70</td>
<td>8.94</td>
<td></td>
</tr>
<tr>
<td>7.1.1.3/2</td>
<td>Moderate</td>
<td>65.00</td>
<td>65</td>
<td>60</td>
<td>55</td>
<td>70</td>
<td>75</td>
<td>75</td>
<td>7.91</td>
<td></td>
</tr>
<tr>
<td>7.1.1.3/3</td>
<td>Moderate</td>
<td>62.00</td>
<td>65</td>
<td>60</td>
<td>55</td>
<td>60</td>
<td>70</td>
<td>70</td>
<td>5.70</td>
<td></td>
</tr>
<tr>
<td>7.1.1.3/4</td>
<td>Moderate</td>
<td>62.00</td>
<td>65</td>
<td>60</td>
<td>55</td>
<td>60</td>
<td>70</td>
<td>70</td>
<td>5.70</td>
<td></td>
</tr>
<tr>
<td>7.1.1.3/5</td>
<td>Moderate</td>
<td>74.00</td>
<td>85</td>
<td>60</td>
<td>80</td>
<td>75</td>
<td>75</td>
<td>70</td>
<td>9.62</td>
<td></td>
</tr>
</tbody>
</table>

**Unit Cut Score:** 65.00

**Moderate Difficulty:**
- 25 - 49.99 Hard
- 50 - 74.99 Moderate
- 75 - 95 Easy

**Difficulty Rating:** A standard deviation of more than 10 will trigger an alert. Discuss the outliers with the judges who set them to determine why. Change as necessary.
Contrasting Groups Method

• This is an examinee-based method
• This method requires that the panel of judges be highly familiar with the test population
• The judges must identify a group of examinees that are clearly non-masters and another group that are clearly masters
• In this method minimally competent or borderline examinees are not included
Contrasting Groups Method (cont.)

• When using this method it is extremely important to select the correct population that is considered to be non-masters
  • Even though they are not minimally competent they should at least have some knowledge of the test area and be members of the target test population
    • Think of a new class reporting for instruction on day one of a course – they know something but not enough to be considered minimally competent

• Both groups are given the test, their test score frequency distributions are plotted on the same continuum, one “line” for masters and one “line” for non-masters

• The point at which the plots intersect is set as the cut or passing score

(Professional Testing, Inc., 2016)
Contrasting Groups Method
Achievable Passing Score

• Once we have established what percentage of minimally competent performers are expected to answer knowledge type test items correctly for particular topics, we can then set achievable standards.

• When these standards are established we can then validate “knowledge gaps” in specific terms.
Identifying the Real Knowledge Gap

• Use reliable analytics
• Make sure all test items are performing well
  • Check “P” and “d” values
• Compare class to class results
  • Identify and correct weak instruction
Using Questionmark® Analytics

Identifying the “real” knowledge gap
Identifying the Knowledge Gap

• Select “Score List” report
  • The Score list report provides ability to view the scores achieved by participants on a particular assessment
  • The Score list report can be run against any non-survey assessment type (i.e., text, exam, or quiz)
    • Provides information by:
      • Assessment results
      • Topic results
      • Item results

<table>
<thead>
<tr>
<th>View</th>
<th>Score list report</th>
<th>Quiz, test, exam results</th>
<th>Provides a sortable list of participants’ scores and aggregate scores for performance comparisons and reporting. Reports assessment, topic, and item-level scores.</th>
</tr>
</thead>
</table>
Assessment Results View

• The aggregate table (assessment) contains five columns:
  • The Group column displays the name of the group
  • The N completed column displays the number of participants with completed results in the group
  • The Avg score column displays the average raw score for the group
  • The Avg % column displays the average percentage score for the group
  • The Score band % distribution column displays the score band distribution for the group
# Aggregate Statistics Table (Assessment)

<table>
<thead>
<tr>
<th>Group</th>
<th>N completed</th>
<th>Avg score</th>
<th>Avg %</th>
<th>Score band % distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Report total</td>
<td>25</td>
<td>4.5</td>
<td>44.8</td>
<td>48 Fail 52 Pass</td>
</tr>
<tr>
<td>Documentation</td>
<td>2</td>
<td>5.0</td>
<td>50.0</td>
<td>50 Fail 50 Pass</td>
</tr>
<tr>
<td>Product Management</td>
<td>2</td>
<td>5.0</td>
<td>50.0</td>
<td>100 Pass</td>
</tr>
<tr>
<td>Quality Assurance</td>
<td>3</td>
<td>3.3</td>
<td>33.3</td>
<td>67 Fail 33 Pass</td>
</tr>
<tr>
<td>Software Engineering</td>
<td>13</td>
<td>4.3</td>
<td>43.1</td>
<td>62 Fail 38 Pass</td>
</tr>
<tr>
<td>Solutions</td>
<td>2</td>
<td>5.5</td>
<td>55.0</td>
<td>100 Pass</td>
</tr>
<tr>
<td>Technical Services</td>
<td>3</td>
<td>5.0</td>
<td>50.0</td>
<td>33 Fail 67 Pass</td>
</tr>
</tbody>
</table>

Score band legend: *Fail* | *Pass*
Topic Results View

• The aggregate statistics table (topic) contains two standard columns:
  • The Group column displays the name of the group to which the participant belongs
  • The N completed column displays the number of participants with completed results in the group

• After these two standard columns, for each topic in the assessment being reported on, two columns will be added to aggregate statistics table
  • One column will be added for the average score on the topic (the header for this column will have the format <Topic name> Avg Score (Max = <Maximum possible score for this topic>))
  • The second column will be added for the average percentage score on the topic (the header for this column will have the format <Topic name> Avg %)
## Aggregate Statistics Table (Topic)

<table>
<thead>
<tr>
<th>Group</th>
<th>N completed</th>
<th>Moder... Avg Score (Max = 5)</th>
<th>Moder... %</th>
<th>Roman... Avg Score (Max = 5)</th>
<th>Roman... %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Report total</td>
<td>25</td>
<td>2.5</td>
<td>49.6</td>
<td>2.0</td>
<td>40.0</td>
</tr>
<tr>
<td>Documentation</td>
<td>2</td>
<td>2.0</td>
<td>40.0</td>
<td>3.0</td>
<td>60.0</td>
</tr>
<tr>
<td>Product Management</td>
<td>2</td>
<td>3.0</td>
<td>60.0</td>
<td>2.0</td>
<td>40.0</td>
</tr>
<tr>
<td>Quality Assurance</td>
<td>3</td>
<td>2.0</td>
<td>40.0</td>
<td>1.3</td>
<td>26.7</td>
</tr>
<tr>
<td>Software Engineering</td>
<td>13</td>
<td>2.3</td>
<td>46.2</td>
<td>2.0</td>
<td>40.0</td>
</tr>
<tr>
<td>Solutions</td>
<td>2</td>
<td>4.0</td>
<td>80.0</td>
<td>1.5</td>
<td>30.0</td>
</tr>
<tr>
<td>Technical Services</td>
<td>3</td>
<td>2.7</td>
<td>53.3</td>
<td>2.3</td>
<td>46.7</td>
</tr>
</tbody>
</table>
The item results view of the report consists of two tables. The first table calculates aggregate statistics, and the second table displays participant results.

The aggregate statistics table contains two standard columns:
- The **Group** column displays the name of the group to which the participant belongs.
- The **N completed** column displays the number of participants with completed results in the group.

Images Copyright Questionmark
Let’s listen in...
So Steve, Julie ran the analytics for the entire Snapperdoodle 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 that the actual test score results are not what we expected. We sorted the results by each topic and workgroup to see what we could find.

So which workgroup isn’t performing? Who gets fired first?
Nobody has to be fired Steve, we just need to find the root cause of the poor average scores of two topics, Snapperdoodle Troubleshooting and Snapperdoodle Error Trapping. There may just be a problem in the curriculum.

So what you’re telling me Stacy, is it may be our fault because we are teaching them the wrong thing?
That is a strong possibility Steve. The average scores on the other eight topics are pretty good. We even sorted the scores by instructor and there was not much difference. I really think we have to look at what we are actually teaching to make sure it matches what we expect them to be doing on the job.

Wow! Maybe it’s my fault! I approved the curriculum. I guess I’ll have to take a closer look at those two topics to make sure we are actually teaching what we should be.
Sure thing Steve! I’m glad we were able to catch this before we got too far into the Snapperdoodle development.

Thanks for looking into this and pointing me in the right direction Stacy! I didn’t know that Questionmark analytics could do that! Be sure to thank Julie for her hard work too!
Good news and bad news Eric. Stacy and Julie have identified where the problem is using Questionmark analytics. Actually it’s me...I had the instructors teaching the wrong thing for two areas. I’m going to review the curriculum tonight and have it correct by tomorrow.
Hi Julie. Great job on identifying the knowledge gap. Steve is going to take a look at the two weak areas and make any changes necessary to the curriculum to make sure it matches what we expect on the job. Steve sends his thanks!

I’m glad I could help Stacy! Good thing we caught this before it got too bad. I’m going to run a complete item analysis of the entire bank to make sure the items are performing well. I’ll let you know what I find.
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
Sample Topic Structure
Future Reporting

• 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
• Unit by Unit
• Unit by sub-topic
KEEP CALM AND MIND THE GAP
Questions?
Website: www.gocompassconsultants.com

E-mail: james.parry@gocompassconsultants.com