This paper is based on a popular series of posts about Item Analysis Analytics from the Questionmark blog. It provides instruction on the principles of item analysis and how to apply them using the Questionmark™ Perception™ assessment management system.

Author: Greg Pope
## Table of contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>3</td>
</tr>
<tr>
<td>Part 1: What is Classical Test Theory?</td>
<td>4</td>
</tr>
<tr>
<td>Part 2: Conducting an Item Analysis</td>
<td>5</td>
</tr>
<tr>
<td>Part 4: The Nitty-Gritty of Item Analysis</td>
<td>7</td>
</tr>
<tr>
<td>Part 5: Outcome Discrimination and Outcome Correlation</td>
<td>10</td>
</tr>
<tr>
<td>Part 6: Determining Whether a Question Makes the Grade</td>
<td>11</td>
</tr>
<tr>
<td>Part 7: The Psychometric Good, Bad and Ugly</td>
<td>13</td>
</tr>
<tr>
<td>Part 8: Some Problematic Questions</td>
<td>16</td>
</tr>
</tbody>
</table>
Introduction

Item analysis is a hot topic for social conversation (Okay, maybe just for some people). I thought it might be useful to talk about Classical Test Theory (CTT) and item analysis analytics in a series of blog posts on the Questionmark blog (http://blog.questionmark.com) to help “demystify” this commonly used approach (and sometimes misunderstood) to analyzing and interpreting test results. After many requests from blog readers, I have compiled these posts into the following paper. Enjoy!
Part 1: What is Classical Test Theory?

Classical Test Theory (CTT) is a body of theory and research regarding psychological testing that predicts/explains the difficulty of questions, provides insight into the reliability of assessment scores, and helps us represent what examinees know and can do. In a similar manner to theories regarding weather prediction or ocean current flow, CTT provides a theoretical framework for understanding educational and psychological measurement. The essential basis of CTT is that many questions combine to produce a measurement (assessment score) representing what a test taker knows and can do.

CTT has been around a long time (since the early 20th century) and is probably the most widely used theory in the area of educational and psychological testing. CTT works well for most assessment applications for reasons such as its ability to work with smaller sample sizes (e.g., 100 or less), and that it is relatively simple to compute and understand the statistics.

The general CTT model is based on the notion that the observed score that test takers obtain from assessments is composed of a theoretical un-measurable “true score” and error. Just as most measurement devices have some error inherent in their measurement (e.g., a thermometer may be accurate to within 0.1 degree 9 times out of 10), so too do assessment scores. For example, if a participant's observed score (what they got reported back to them) on an exam was 86%, their “true score” may actually be between 80% and 92%.

Measurement error can be estimated and relates back to reliability: greater assessment score reliability means less error of measurement. Why does error relate so directly to reliability? Well, reliability has to do with measurement consistency. So if you could take the average of all the scores that a participant obtained—if they took the same assessment an infinite number of times with no remembering effects—this would be a participant’s true score. The more reliability in the measurement the less wildly diverse the scores would be each time a participant took that assessment over eternity.

For a more detailed overview of CTT, that won’t make your lobes fall off, try Chapter 5 in Dr. Theresa Kline's book, "Psychological Testing: A Practical Approach to Design and Evaluation."
Part 2: Conducting an Item Analysis

Classical Test Theory (CTT) provides a foundation for discussing item analysis analytics using CTT. Now let’s talk about the high-level purpose and process of conducting an item analysis. The general purpose of conducting an item analysis is to find out whether the questions composing an assessment are performing in a manner that is psychometrically appropriate and defensible. Item analyses are used to evaluate the psychometric performance of questions. They help us find out whether items need to be improved (sent back to development), sent to the scrap heap, or left as they are because they meet all the criteria for being included in an assessment.

I’d like to share a tip about how some of my colleagues decide whether to revise a problematic looking question or throw it away as “unfixable.” This involves setting a review time limit for each question that needs to be reviewed. In an item analysis review meeting, this may involve:

- Psychometricians
- Subject Matter Experts
- Exam developers
- Other stakeholders

Each question could be reviewed for no more than a pre-determined period of time, say 10 minutes. If an effective revision for the question does not become apparent within that period of time, the question goes to the scrap bin and a new question is developed by SMEs to take its place.

Many organizations beta test questions in order to choose those that should be included in an actual assessment. Questionmark Perception offers the delivery status field of “Experimental.” This allows beta questions to be included/interspersed within an actual assessment form but not scored and therefore not counted as part of the calculation of participant assessment scores.

For more information about beta testing questions using Perception, please refer to the following Best Practice Guide: Beta Testing High-stakes Questions with Questionmark Perception.
Part 3: What to Look for in an Item Analysis Report

Here are some of the essential things to look for in a typical Item Analysis Report:

- **Number of results**
  - How many participants have answered the question
  - Important to know how stable the statistics are, generally more participants answering the question means more stability of statistics

- **Number not answering**
  - How many participants have not answered the question
  - Important to know if there was a problem with the question wording (participants got confused and didn’t answer) or the display (an image wasn’t displaying properly on screen) that need to be addressed
  - Also impacts the number of results

- **Question difficulty (p-value)**
  - The difficulty of a question represented as a proportion of the participants that answered the question correctly (in the case of multiple choice items)
  - A p-value for a question of 0.550 means that 55% of participants choose the correct answer for the question

- **Question discrimination**
  - The degree to which questions differentiate or discriminate between participants who know the material well and those that do not know the material well. Examples of discrimination statistics:
    - **High minus Low**: The proportion of examinees of the top 27% of participants (in terms of assessment scores) minus the lowest 27% (should be large positive difference between high and low)
    - **Item-total correlation coefficient**: A Point-biserial correlation between examinee assessment scores and item scores (higher item scores should mean higher assessment scores)

You may sometimes see “Alpha if item deleted” statistics in Item Analysis Reports. These statistics provide information about whether the internal consistency reliability (e.g., Cronbach’s Alpha) will increase if the question is deleted from the assessment. An increase in the reliability value indicates that the question is not performing well psychometrically. Many Item Analysis Reports do not display the “Alpha if item deleted” statistic because the item-total correlation coefficient provides basically the same information. Questions with:

- Higher item-total correlation coefficient values will contribute to higher internal consistency reliability values
- Lower item-total correlation coefficient values will contribute to lower internal consistency reliability values.

Other statistics you might see are variations of the point-biserial item-total correlation coefficient such as “Corrected Point-biserial correlation,” “biserial correlation” or “corrected biserial correlation.” The “corrected” in these refers to taking out the question scores from the calculations so that the question being examined is not “contributing to itself” in terms of the statistics.

A great resource for more information on item analysis is Chapter 8 of Dr. Steven J. Osterlind’s book *Constructing Test Items: Multiple-Choice, Constructed-Response, Performance and Other Formats (2nd edition).*
Part 4: The Nitty-Gritty of Item Analysis

Let’s examine some example questions and explain how to use the Questionmark Item Analysis Report in an applied context for a State Capitals Exam.

The Questionmark Item Analysis Report first produces an overview of question performance both in terms of the difficulty of questions and in terms of the discrimination of questions (upper minus lower groups). These overview charts give you a “bird’s eye view” of how the questions composing an assessment perform. In the example below we see that we have a range of questions in terms of their difficulty (“Item Difficulty Level Histogram”), with some harder questions (the bars on the left), most average-difficulty questions (bars in the middle), and some easier questions (the bars on the right). In terms of discrimination (“Discrimination Indices Histogram”) we see that we have many questions that have high discrimination as evidenced by the bars being pushed up to the right (more questions on the assessment have higher discrimination statistics).

Overall, if I were building a typical criterion-referenced assessment with a pass score around 50% I would be quite happy with this picture. We have more questions functioning at the pass score point with a range of questions surrounding it and lots of highly discriminating questions. We do have one rogue question on the far left with a very low discrimination index, which we need to look at.

The next step is to drill down into each question to ensure that each question performs as it should. Let’s look at two questions from this assessment, one question that performs well and one question that does not perform so well.

The question below is an example of a question that performs nicely. Here are some reasons why:

- Going from left to right, first we see that the “Number of Results” is 175, which is a nice sample of participants to evaluate the psychometric performance of this question.
- Next we see that everyone answered the question (“Number not Answered” = 0), which means there probably wasn’t a problem with people not finishing or finding the questions confusing and giving up.
The “P Value Proportion Correct” shows us that this question is just above the pass score where 61% of participants ‘got it right.’ There is nothing wrong with that: the question is neither too easy nor too hard.

The “Item Discrimination” indicates good discrimination, with the difference between the upper and lower group in terms of the proportion selecting the correct answer of ‘Salem’ at 48%. This means that of the participants with high overall exam scores, 88% selected the correct answer versus only 40% of the participants with the lowest overall exam scores. This is a nice, expected pattern.

The “Item Total Correlation” backs the Item Discrimination up with a strong value of 0.40. This means that of all participants who answered the questions, the pattern of high scorers getting the question right more than low scorers holds true.

Finally we look at the Outcome information to see how the distracters perform. We find that each distracter pulled some participants, with ‘Portland’ pulling the most participants, especially from the “Lower Group.” This pattern makes sense because those with poor state capital knowledge may make the common mistake of selecting Portland as the capital of Oregon.

The psychometricians, SMEs, and test developers reviewing this question all have smiles on their faces when they see the item analysis for this item.

The rogue question that does not perform so well in terms of discrimination—the one we saw in the Discrimination Indices Histogram. When we look into the question we understand why it was flagged:

- Going from left to right, first we see that the “Number of Results” is 175, which is again a nice sample size: nothing wrong here.
- Next we see everyone answered the question, which is good.
- The first red flag comes from the “P Value Proportion Correct” as this question is quite difficult (only 35% of participants selected the correct answer). This is not in and of itself a bad thing so we can keep this in memory as we move on,
- The “Item Discrimination” indicates a major problem, a negative discrimination value. This means that participants with the lowest exam scores selected the correct answer more than participants with the highest exam scores. This is not the expected pattern we are looking for: Houston, this question has a problem!
- The “Item Total Correlation” backs up the Item Discrimination with a high negative value.
- To find out more about what is going on we delve into the Outcome information area to see how the distracters perform. We find that the keyed-correct answer of Nampa is not showing the expected pattern of upper minus lower proportions. We do, however, find that the distracter “Boise” is showing the expected pattern of the Upper Group (86%) selecting this.
response option much more than the Lower Group (15%). Wait a second...I think I know what is wrong with this one, it has been mis-keyed! Someone accidentally assigned a score of 1 to Nampa rather than Boise.

No problem: the administrator pulls the data into the Results Management System (RMS), changes the keyed correct answer to Boise, and presto, we now have defensible statistics that we can work with for this question.

The psychometricians, SMEs, and test developers reviewing this question had a frown on their faces at first but those frowns were turned upside down when they realized it is just a simple mis-keyed question.
Part 5: Outcome Discrimination and Outcome Correlation

People often get confused about outcome discrimination and outcome correlation. How are these different or the same? When should you use one or the other? The fact of the matter is that you can use one or the other and often it comes down to preference as they both yield quite similar results.

Outcome discrimination is the proportion of the top (27% according to assessment score) of participants who selected a response option minus the lowest (27% according to assessment score) of participants who selected each response option to the question. What you would expect is that participants with the highest assessment scores should select the correct response option more often than participants with the lowest assessment scores. Similarly, participants with the highest assessment scores should select the incorrect distracters less often compared to the participants with the lowest assessment scores.

Outcome correlation is a point-biserial correlation that correlates the outcomes scores that participants achieve to the assessment scores that they achieve. So rather than comparing only the top and bottom 27% of participants, the outcome correlation looks at all participants using a standard correlation approach.

If you are thinking that outcome discrimination and outcome correlation sound like they might be related to one another, you are right! High outcome discrimination statistics generally will result in high outcome correlations. In other words, outcome discrimination and outcome correlation statistics are highly correlated with one another. How correlated are they? Well, I looked at many real-life questions from Item Analysis Reports that Questionmark customers have shared with me and found a positive correlation of 0.962, which is really high.
Part 6: Determining Whether a Question Makes the Grade

Here are some criteria that can be used for outcome discrimination and outcome correlation coefficients to judge whether a question is making the grade in terms of psychometric quality.

Outcome discrimination (high-low)

- **Negative**
  - Major problem if this is occurring for the correct response option (Intervention! Find out why!)
  - This is expected for the incorrect distracters

- **0 to 0.15**
  - Low outcome discrimination
  - These questions should be reviewed to determine why

- **0.16 to 0.29**
  - Moderate outcome discrimination

- **0.30 to 0.49**
  - High outcome discrimination

- **0.5 or greater**
  - Very high outcome discrimination
### Outcome correlation (Point-biserial correlation)

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
</table>
| **Negative**      | • Major problem if this is occurring for the correct response option (Intervention! Find out why!)  
                     • This is expected for the incorrect distractors |
| **Around 0**      | • No relationship between the question score and the assessment score  
                     • These questions need to be reviewed to determine what is happening |
| **0 to 0.19**     | • Low correlation between outcome scores and assessment scores |
| **0.2 to 0.29**   | • Moderate correlation between outcome scores and assessment scores |
| **0.3 to 0.44**   | • Strong correlation between outcome scores and assessment scores |
| **0.45 or greater** | • Very strong correlation between outcome scores and assessment scores |
Part 7: The Psychometric Good, Bad and Ugly

Part 4 of this paper showed an example item analysis report for a question that performed well statistically and a question that did not perform well statistically. The latter turned out to be a mis-keyed item. It’s interesting to drill into a few more item analysis cases of questions that have interesting psychometric performance. This should help people recognize the patterns of the psychometric good, bad and ugly in terms of question performance.

The question below is an example of a question that is borderline in terms of psychometric performance. Here are some reasons why:

- Going from left to right, first we see that the “Number of Results” is 116, which is a decent sample of participants to evaluate the psychometric performance of this question.
- Next we see everyone answered the question (“Number not Answered” = 0) which means there probably wasn’t a problem with people not finishing or finding the questions confusing and giving up.
- The “P Value Proportion Correct” shows us that this question is average to easy, with 65% of participants “getting it right.”
- The “Item Discrimination” indicates mediocre discrimination at best, with the difference between the upper and lower group in terms of the proportion selecting the correct answer of ‘Leptokurtic’ at 20%. This means that of the participants with high overall exam scores, 75% selected the correct answer versus 55% of the participants with the lowest overall exam scores. I would have liked to see a larger difference between the Upper and Lower groups.
- The “Item Total Correlation” backs the Item Discrimination up with a lackluster value of 0.14. A value like this would likely not meet many organizations’ internal criteria in terms of what is considered a “good” item.
- Finally, we look at the Outcome information to see how the distracters perform. We find that each distracter pulls some participants, with ‘Platykurtic’ pulling the most participants and quite a large number of the Upper group (22%) selecting this distracter. If I were to guess what is happening, I would say that because the correct option and the distracters are so similar, and because this topic is so obscure you really need to know your material, participants get confused between the correct answer of ‘Leptokurtic’ and the distracter ‘Platykurtic’

The psychometricians, SMEs, and test developers reviewing this question would need to talk with instructors to find out more about how this topic was taught and understand where the problem lies: Is it a problem with the question wording or a problem with instruction and retention/recall of material? If it is a question wording problem, revisions can be made and the question re-beta tested. If the problem is in how the material is being taught, then instructional coaching can occur and the question re-beta tested as it is to see if improvements in the psychometric performance of the question occur.
The question below is an example of a question that has a classic problem. Here are some reasons why it is problematic:

- Going from left to right, first we see that the “Number of Results” is 175. That is a fairly healthy sample, nothing wrong there.

- Next we see everyone answered the question ("Number not Answered" = 0), which means there probably wasn’t a problem with people not finishing or finding the question confusing and giving up.

- The “P Value Proportion Correct” shows us that this question is easy, with 83% of participants ‘getting it right’. There is nothing immediately wrong with an easy question, so let’s look further.

- The “Item Discrimination” indicates reasonable discrimination, with the difference between the Upper and Lower group in terms of the proportion selecting the correct answer of ‘Cronbach’s Alpha’ at 38%. This means that of the participants with high overall exam scores, 98% selected the correct answer versus 60% of the participants with the lowest overall exam scores. That is a nice difference between the Upper and Lower groups, with almost 100% of the Upper group choosing the correct answer. Obviously, this question is easy for participants who know their stuff!

- The “Item Total Correlation” backs the Item Discrimination up with a value of 0.39. This value backs up the “Item Discrimination” statistics and would meet most organizations’ internal criteria in terms of what is considered a “good” item.

- Finally, we look at the Outcome information to see how the distracters perform. Well, two of the distracters don’t pull any participants! This is a waste of good question real estate: Participants have to read through four alternatives when there are only two they even consider as being the correct answer.

The psychometricians, SMEs, and test developers reviewing this question would likely ask the SME who developed the question to come up with better distracters that would draw more participants. Clearly, ‘Bob’s Alpha’ is a joke distracter that participants dismiss immediately as is the ‘KR-1,000,000’, I mean Kuder-Richardson formula one million. Let’s get serious here!
<table>
<thead>
<tr>
<th>Number of Results</th>
<th>Number not Answered</th>
<th>P Value (Proportion Correct)</th>
<th>Item Discrimination</th>
<th>Item Total Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>175</td>
<td>0</td>
<td>0.83</td>
<td>0.38</td>
<td>0.39</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outcome Name</th>
<th>Times Occurred (Frequency)</th>
<th>Proportion Selected</th>
<th>Upper Group</th>
<th>Lower Group</th>
<th>Outcome Discrimination</th>
<th>Outcome Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bob's Alpha</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>KR-1,000,000</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><em>Cronbach's</em></td>
<td></td>
<td></td>
<td>146</td>
<td>0.83</td>
<td>0.98</td>
<td>0.60</td>
</tr>
<tr>
<td>Alpha-Correct</td>
<td></td>
<td></td>
<td>23</td>
<td>0.16</td>
<td>0.02</td>
<td>0.40</td>
</tr>
<tr>
<td>Not answered</td>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>TOTAL</td>
<td>175</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Question: What is the internal consistency reliability statistic that the Questionmark Test Analysis Report uses?
Part 8: Some Problematic Questions

It might be useful to show a few examples of some questions that have bad and downright terrible psychometric performance to show the ugly side of item analysis.

Below is an example of a question that is fairly terrible in terms psychometric performance. Here are some reasons why:

- Going from left to right, first we see that the “Number of Results” is 65, which is not so good: there are too few participants in the sample to be able to make sound judgments about the psychometric performance of the question.

- Next we see that 25 participants didn’t answer the question (“Number not Answered” = 25), which means there was a problem with people not finishing or finding the questions confusing and giving up.

- The “P Value Proportion Correct” shows us that this question is hard with 20% of participants ‘getting it right.’

- The “Item Discrimination” indicates very low discrimination, with the difference between the Upper and Lower group in terms of the proportion selecting the correct answer of ‘More than 40’ at only 5%. This means that of the participants with high overall exam scores, 27% selected the correct answer versus 22% of the participants with the lowest overall exam scores. This is a very small difference between the Upper and Lower groups. Participants who know the material should have got the question right more often.

- The “Item Total Correlation” reflects the Item Discrimination with a negative value of -0.01. A value like this would definitely not meet most organizations’ internal criteria in terms of what is considered an acceptable item. Negative item-total correlations are a major red flag!

- Finally we look at the Outcome information to see how the distracters perform. We find that participants are all over the map selecting distracters in an erratic way. When I look at the question wording I realize how vague and arbitrary this question is: the number of questions that should be in an assessment depends on numerous factors and contexts. It is impossible to say that in any context a certain number of questions are required. It looks like the Upper Group are selecting the response options ‘21-40’ and ‘More than 40’ response options more than the other two options, which have smaller numbers of questions. This makes sense from a participant guessing perspective, because in many assessment contexts having more questions than fewer questions is better for reliability.

The psychometricians, SMEs, and test developers reviewing this question would need to send the SME who wrote this question back to basic authoring training to ensure that they know how to write questions that are clear and concise. This question does not really have a correct answer and needs to be re-written to clarify the context and provide many more details to the participants. I would even be tempted to throw out questions along this content line, because how long an assessment should be has no one “right answer.” How long an assessment should be depends on so many things that there will always be room for ambiguity, so it would be quite challenging to write a question that performs well statistically on this topic.
Below is an example of a question that is downright awful in terms psychometric performance. Here are some reasons why:

- Going from left to right, first we see that the “Number of Results” is 268, which is really good. That is a nice healthy sample. Nothing wrong here, let’s move on.

- Next we see that 56 participants didn’t answer the question (“Number not Answered” = 56), which means there was a problem with people not finishing or finding the questions confusing and giving up. It gets worse, much, much worse.

- The “P Value Proportion Correct” shows us that this question is really hard, with 16% of participants ‘getting it right.’

- The “Item Discrimination” indicates a negative discrimination, with the difference between the Upper and Lower group in terms of the proportion selecting the correct answer of ‘44123’ at minus 23%. This means that of the participants with high overall exam scores, only 12% selected the correct answer versus 35% of the participants with the lowest overall exam scores. What the heck is going on? This means that participants with the highest overall assessment scores are selecting the correct answer LESS OFTEN than participants with the lowest overall assessment scores. That is not good at all; let’s dig deeper.

- The “Item Total Correlation” reflects the Item Discrimination with a large negative value of minus 0.26. This is a clear indication that there is something incredibly wrong with this question.

- Finally we look at the Outcome information to see how the distracters perform. This is where the true psychometric horror of this question is manifested. There is neither rhyme nor reason here: participants, regardless of their performance on the overall assessment, are all over the place in terms of selecting response options. You might as well have blindfolded everyone taking this question and had them randomly select their answers. This must have been extremely frustrating for the participants who had to take this question and would have likely led to many participants thinking that the organization administering this question did not know what they were doing.

The psychometricians, SMEs, and test developers reviewing this question would need to provide a pink slip to the SME who wrote this question immediately. Clearly the SME failed basic question authoring training. This question makes no sense and was written in such a way to suggest that the author was under the influence, or otherwise not in a right state of mind, when crafting this
question. What is this question testing? How can anyone possibly make sense of this and come up with a correct answer? Is there a correct answer? This question is not salvageable and should be stricken from the Perception repository without a second thought. A question like this should have never gotten in front of a participant to take, let alone 268 participants. The panel reviewing questions should review their processes to ensure that in the future questions like this are weeded out before an assessment goes out live for people to take.

<table>
<thead>
<tr>
<th>Number of Results</th>
<th>Number not Answered</th>
<th>P Value Proportion Correct</th>
<th>Item Discrimination</th>
<th>Item Total Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>268</td>
<td>56</td>
<td>0.16</td>
<td>-0.23</td>
<td>-0.25</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outcome Name</th>
<th>Times Occurred (Frequency)</th>
<th>Proportion Selected</th>
<th>Upper Group</th>
<th>Lower Group</th>
<th>Outcome Discrimination</th>
<th>Outcome Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>34.12</td>
<td>54</td>
<td>0.20</td>
<td>0.22</td>
<td>0.26</td>
<td>-0.04</td>
<td>-0.09</td>
</tr>
<tr>
<td>84.12&lt;Correct&gt;</td>
<td>42</td>
<td>0.16</td>
<td>0.12</td>
<td>0.35</td>
<td>-0.23</td>
<td>-0.26</td>
</tr>
<tr>
<td>1.01</td>
<td>51</td>
<td>0.19</td>
<td>0.23</td>
<td>0.25</td>
<td>-0.02</td>
<td>-0.07</td>
</tr>
<tr>
<td>7</td>
<td>65</td>
<td>0.24</td>
<td>0.19</td>
<td>0.14</td>
<td>0.05</td>
<td>-0.01</td>
</tr>
<tr>
<td>Not answered</td>
<td>56</td>
<td>0.21</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>TOTAL</td>
<td>268</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Question – What is the difference if you multiple three plus two and then subtract the division of seven digits?
About Questionmark:

Questionmark provides technologies and services that enable organizations to measure knowledge, skills and attitudes securely for certification, regulatory compliance and improved learning outcomes. Questionmark solutions enable reliable, valid and defensible assessments by empowering subject matter experts through collaborative authoring, accommodating participant needs with blended and multilingual delivery and informing stakeholders through timely reporting and meaningful analytics.

The Questionmark Perception assessment management system, available as a licensed, hosted or subscription-based solution, enables organizations to create, administer and report on surveys, quizzes, tests and exams. Complete details are available at http://www.questionmark.com