Attribute Agreement Analysis

By Arvind Srivastava
MBB, CQA, IATF Lead Auditor, ASQ William Lieberman Honoree
Corporate MBB/IE, Accuride Corporation
Learning Objectives

Across this time we will…

• Learn about how to Calibrate Humans via Attribute Agreement Analysis (AAA) to mitigate business loss
• Understand the basic procedures involved in conducting an AAA to objectively qualify a “human gage”
• Build confidence in how to run an AAA via simulation
• Evaluate AAA deployment across the supply chain
• Review examples of AAA impact
An Attribute MSA

Humans believe what they see. We do not generally question our expertise when assessing good or bad. There is generally **one type** of MSA of assessment conducted for Attributes to determine **how** right or wrong we are in our results. This is an Attributes Agreement Analysis (AAA). It is used for visual assessment, go/no go testing or when results are being categorized.
Measurement System Analysis
aka MSA

When we measure or make an assessment of the goodness of an item, we need to be sure that our result is correct. If it is not correct, we take two risks:

- Alpha $\alpha$ Risk: We may inadvertently discard or rework a good item (Aw, darn)
- Beta $\beta$ Risk: We may inadvertently pass on a bad item (Boy, that was Bad)
## Alpha/Beta Impacts

What are the typical impacts of either type of rejection scenario:

<table>
<thead>
<tr>
<th>Location</th>
<th>Alpha $\alpha$ (reject good)</th>
<th>Beta $\beta$ (pass bad)</th>
</tr>
</thead>
</table>
| **Internal to the company** | • Unnecessary scrap  
• Material loss  
• Unnecessary rework  
• General margin loss  
• Arguments…who’s right? | • Customer reject/ill will  
• Containment expenses  
• Increased inspection criteria  
• Unplanned cost increase |
| **Upstream Supplier to the company** | • Same as above  
• Increased product cost to customer to manage internal “losses” | • Same as above |
| **Downstream Customer** | • Unintended increase in product standard  
• Unintended cost increase of product (margin erosion) | • Same as above  
• Business risk depending on supplier scoring program  
• DPPM is DPPM is DPPM |
Why Do We Need To Know?

We need to know how much error there is in our measurement processes for several reasons:

- Prevent $\alpha$ and $\beta$ errors; Reducing scrap/rework
- Reduce customer arguments over good/bad
- It is our **JOB** to ensure that our people are enabled to make the right pass/fail decision **every** time
- MSA is required by **IATF/AS/TL**…as a part of PPAP

**Note**: In those advanced systems every line item on the control plan called out for evaluation is **required** to have an MSA conducted. Some think that this only applies to variables measurement
Risk Assessment

How might this tie to Risk-Based Thinking in the newly revised ISO 9001:2015?

• While **not** called out formally in ISO 9001:2015, MSA is called out in many secondary compliance documents such as IATF, AS, TL, etc.

• AAA is a form of MSA for attributes. Most people generally conduct only variables gaging MSA’s called “Gage R&R” or GRR

• Consider inclusion in PFMEA/CP documentation via the APQP (or commercialization) process
Let’s $Play $Ping $Pong!

Even though the business is for a safety product with performance critical dimensions, almost all product failures were for attribute reasons: Paint Color/Shade, Gloss/Shininess, Surface Defects, Scratches, etc.

• Several sites made the same PNs: There was inconsistency both across and within the various locations
• Internally, personnel were inconsistent on what was or wasn’t good; over-inspection resulted. This inadvertently increased the finished goods expectation (beyond the spec) where in-spec product was now no longer acceptable
• Product would be shipped back and forth for rework and/or replacement; a very expensive game of ping pong. Let’s change the game!
When attribute based rejections occur across the value stream, they can be very expensive and time-consuming.
Attribute Agreement Analysis

AAA Checks for the chances of 100% agreement on three features:

- **Within** “myself”; Did I repeatedly call it good or bad in a consistent manner (even if I was wrong)

- **Between** me and a “peer”; Did both my peer and I repeatedly call it good or bad in a consistent manner (even if we were both wrong)

- Compared to “Standard”; Did I/we get it right each time

<table>
<thead>
<tr>
<th></th>
<th>Std</th>
<th>Op1</th>
<th>Op2</th>
<th>Op3</th>
<th>Y/N w/in</th>
<th>Y/N Std</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>2</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>F</td>
<td>F</td>
<td>Y</td>
</tr>
<tr>
<td>3</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>P</td>
<td>P</td>
<td>N</td>
</tr>
<tr>
<td>4</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>5</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>Y</td>
</tr>
<tr>
<td>6</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>Y</td>
</tr>
</tbody>
</table>
AAA Needs

For an Attributes Agreement Analysis (AAA), two things are needed:

➢ The Pass/Fail “Samples”: Preferably 50 or more.  
   **Note**: One unit may have multiple samples on it

➢ The AAA: A check for accuracy in human performance between the appraisers and the standard. A typical target for “Statistical Agreement” is >= 85%

**Note**: The Statistical Agreement can be modified by the amount of agreement that happens by chance. This evaluation is known as Kappa (K) statistics. The AIAG identifies K >= 75%. Minitab provides both formats.
AAA: What It Looks Like

The AAA Gives a series of graphs to show how the assessors generally perform. The goal is to be in 100% agreement. While not feasible, 85% Statistical Agreement is usually acceptable. (75% Kappa $^{\text{AIAG}}$)

Not an effective Statistical Agreement at < 85%... This team agrees with each other and the standard about 68% of the time. However, 95% of the time, they will range in agreement from 47% to 85%
AAA: Triggering Exercise

BEFORE the process “hits the floor”, an AAA should be evaluated like a GRR for its effectiveness (in the PFMEA/CP). In the gaging system, AAAs should be performed like “PM” to keep assessors “calibrated”. When should an AAA be conducted?

<table>
<thead>
<tr>
<th>Man</th>
<th>Machine</th>
<th>Material</th>
</tr>
</thead>
<tbody>
<tr>
<td>New operator</td>
<td>Equipment changes</td>
<td>Supplier change</td>
</tr>
<tr>
<td>Requalification</td>
<td>New fixtures/dies</td>
<td>Lot qualification</td>
</tr>
<tr>
<td>Change in procedure</td>
<td>New standards</td>
<td>Change in lighting</td>
</tr>
<tr>
<td>Process CI</td>
<td>Customer calibration</td>
<td>Across differing locales</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ambient sensitivities</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Triggers of when to conduct an AAA
Kappa (K) Statistic: The degree of Absolute agreement between ratings. This is NOT the same as Statistical Agreement and is generally done in a statistical tool such as Minitab, JMP, etc.

Attribute decisions are not usually “life or death” but more often of a cosmetic nature. Scratches, pits, color, etc. They are evaluated for “okay-ness” and then passed along.

As such, the “vanilla” form of the “Statistical Agreement” AAA is typically applied. The vanilla version can also be easily done in Excel. For a more critical AAA evaluation, consider the Kappa (K) adjusted AAA.
AAA: Vanilla or Hot Fudge Sundae

𔄐 **AAA in Minitab**: Minitab runs both Statistical agreement and Absolute formats. Be sure to know which one you are running. The version in Assistant (v17) is Statistical

𔄐 **Kendall’s Coefficient**: If the inspection data is ordinal (Ex. Good/better/best rating, 1-5 rating, etc.)
AAA in Excel: Via ASQ Stats Div

**Excel:**
- ASQ Statistics Division issued the format at right to SD members several years ago. It is an easy Excel file to enter data into. The graphs and results automatically populate.
- Purchase gaging software

**Steps:**
- Enter operator data
- Evaluate results
- Apply necessary actions
- This format provides %Agreement (≥ 85%), not \( K \) (≥ 75%)

<table>
<thead>
<tr>
<th>Source</th>
<th>% Appraiser 1</th>
<th>% Appraiser 2</th>
<th>% Appraiser 3</th>
<th>% Score vs Attribute 1</th>
<th>% Score vs Attribute 2</th>
<th>% Score vs Attribute 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Inspected</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td># Matched</td>
<td>14</td>
<td>11</td>
<td>14</td>
<td>11</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>False Negative (operator rejected good product)</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>False Positive (operator accepted bad product)</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mixed</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>95% UCL</td>
<td>100.0%</td>
<td>95.3%</td>
<td>100.0%</td>
<td>95.3%</td>
<td>87.2%</td>
<td>91.6%</td>
</tr>
<tr>
<td>Calculated Score</td>
<td>100.0%</td>
<td>78.6%</td>
<td>100.0%</td>
<td>78.6%</td>
<td>64.3%</td>
<td>71.4%</td>
</tr>
<tr>
<td>95% LCL</td>
<td>76.8%</td>
<td>49.2%</td>
<td>76.8%</td>
<td>49.2%</td>
<td>35.1%</td>
<td>41.9%</td>
</tr>
</tbody>
</table>

**Notes:**
1. Operator agrees with him/herself on both trials
2. Operator agrees on both trials with the known standard
3. All operators agreed within and between themselves
4. All operators agreed within & between themselves AND agreed with the known standard
How Easy Is It To Do?

- Let’s play with pennies
- Assess the pennies for pass/fail
- Enter the data
- Evaluate the process
- Do it again
- One person records the data
- Is the penny “shiny”? The answer is Yes or No
- 3 “operators” assess for “Yes” or “No”
- If Yes, write “Y”
- If No, write “N”
- Conduct 2 rounds with 3 operators
- Visually review results

<table>
<thead>
<tr>
<th>#</th>
<th>1-1st</th>
<th>1-2nd</th>
<th>2-1st</th>
<th>2-2nd</th>
<th>3-1st</th>
<th>3-2nd</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
AAA Across Logistics

Across ~9 mos, 3 countries and 4 Customers…

Site 1
AAA: Mgr
AAA: Insp
AAA: Ops
AAA: Eng

Site 2
AAA: Mgr
AAA: Insp
AAA: Ops
AAA: Eng

Site 3
AAA: Mgr
AAA: Insp
AAA: Ops
AAA: Eng

AAA: HQ

AAA: OSPs

AAA: Customer 1
AAA: Customer 2
AAA: Customer 3
AAA: Customer 4

Paddles are down…
Few to no returns
Example magnitude of progression. All teams demonstrated improved repeatability over time. Some external customers discontinued the attribute evaluation by converting over to a variables form of analysis.
The Game has Changed

• All sites creating the same product went through 3 phases of AAA evaluation levels: Operator, Inspector, and site Management
• They then shared the product samples across sister sites to refine internal inspection definitions
• The sister sites along with engineering then all aligned on the product approval definitions (photos, videos, samples)
• Initially, a key customer was initially partnered with. Together the attribute product specs were re-verified. Site teams deployed to the customer locales to train them on the samples

1. Internal rejects down by >30% across the business
2. External rejects down by 60-80% for key customers
3. Premium freight down across the business
Switch Hitting…

• This entire scenario, while real, was after the fact
• All of those years of freight, alpha/beta rejects, ill-will, etc. could have and should have been prevented
• There was an inadvertent increase in customer expectation due to heightened inspection (alpha errors due to over rejection). The original level was not able to be returned to creating a costlier product
• This is a key form of risk management; at many levels
• While these are visual examples, this tool can be used for taste (food), smell, hearing and touch

How can we be more proactive when evaluating attribute features?
Learning Objectives
Hang Up the Paddles…

Across this time we have…

• Learned how to Calibrate Humans via Attribute Agreement Analysis (AAA)
• Worked to understand the basic procedures involved in conducting an AAA
• Reviewed an example of AAA impact

• asrivast@AccurideCorp.com
• 815-218-8503