Check inspection agreement and escape risk

Read a reference-classification matrix, distinguish overall agreement from the two error directions, and recognize the limits of a single pass.

High agreement can hide an important error direction

Fictional case: inspector Leo assesses 40 reference items. Of 20 reference-bad items,18 are rejected and 2 accepted. Of 20 reference-good items,19 are accepted and 1 rejected. The core matrix retains these exact counts. The team calls 92.5% agreement proof that inspection is repeatable.

Your role and the reference condition

Inspection-method owner and appraiser coach
Normal: Reference status is trustworthy for the defined characteristic; item identity and decision are retained; repeated and between-appraiser comparisons are designed separately.
Gap: One aggregate percentage hides false acceptance and false rejection, and one pass has no repeated classifications from which to judge repeatability.

Work within these conditions

1. This is an illustrative single pass, not a completed attribute measurement-system qualification.

2. The balanced 20/20 reference sample is not a production prevalence estimate.

Fictional teaching case. Supplied observations support the exercise; proposed actions are not verified customer results.

See how the method works

Overall agreement can conceal the error that matters most. Establish a defensible reference classification and include representative, difficult items. A full attribute study uses blinded repeats and multiple appraisers to examine consistency within people, between people and against the reference. The single matrix here illustrates denominator choices only. False acceptance asks what fraction of reference-bad items escaped; false rejection asks what fraction of reference-good items were rejected. Both differ from overall agreement. Report the counts and uncertainty rather than declaring success from one percentage. A reference itself can be uncertain, so unresolved classifications need an explicit adjudication method.

	Classified reject	Classified accept
Reference bad	**18 items**	**2 items**
Reference good	**1 item**	**19 items**

Overall agreement: 37/40 = 92.5%

False acceptance among bad: 2/20 = 10%

False rejection among good: 1/20 = 5%

One illustrative pass only. Within-appraiser repeatability requires blinded repeated classifications.

One illustrative pass only. Within-appraiser repeatability requires blinded repeated classifications.

Follow the evidence and decisions

Inspect the supplied case inputs

Reference class	Rejected	Accepted
Bad	18	2
Good	1	19

01 / Verify reference and observation identity

Leo's coach confirms the definition of good/bad and how the reference decisions were established. Each item keeps one identity linked to Leo's classification.

Why: Agreement against a questionable reference does not establish correctness. The comparison must concern the same characteristic and decision rule.

Evidence / check: Forty item records reconcile to the two reference groups.

02 / Read both correct cells

Correct classifications are 18 bad rejected plus 19 good accepted, totaling 37. Overall agreement is 37/40=92.5%.

Why: Agreement counts matches to the reference across both classes. It says nothing yet about consistency on a repeated blind pass.

Evidence / check: The reported numerator is 37, not merely the count of accepted items.

03 / Separate false acceptance

Two of the 20 reference-bad items were accepted, giving 10% false acceptance among bad items in this study.

Why: The denominator is the reference-bad group. Dividing by all 40 would answer a different question and conceal the conditional risk direction.

Evidence / check: The record identifies the two item IDs for review without claiming a production escape rate.

04 / Separate false rejection

One of 20 reference-good items was rejected, giving 5% false rejection among good items. The coach preserves this distinction from false acceptance.

Why: Both errors matter but can have different consequences. A single combined error percentage can obscure what instruction or decision boundary needs attention.

Evidence / check: The review retains separate error directions and their respective denominators.

05 / Design the next evidence

The coach plans blinded repeat classifications and, where appropriate, other appraisers using representative cases including difficult boundaries. Prior answers are not shown during repeat assessment.

Why: Within-appraiser consistency, between-appraiser agreement and reference agreement are different questions. Repetition must be designed to measure them rather than rehearse remembered labels.

Evidence / check: The next study plan states the comparison and keeps its results blank until performed.

Completed classification interpretation record

Measure	Calculation	What it establishes here
Overall agreement	37/40=92.5%	Single-pass reference matches
False acceptance among bad	2/20=10%	Error direction in selected bad group
False rejection among good	1/20=5%	Error direction in selected good group
Repeatability	No repeated pass supplied	Not established

Reference status is disputed

One item marked reference-bad has conflicting expert assessments and no resolved decision rule.

Decision

Resolve or explicitly classify the reference uncertainty before using that item to score an appraiser's correctness. Preserve the disagreement as useful method evidence.

Reasoning

A forced reference label could turn legitimate ambiguity into an apparent operator failure. This is a method-definition issue as well as a training question.

What would change the decision?

The study record states reference uncertainty and the responsible adjudication process.

Practise with a changed case

A new matrix with different class sizes

New fictional pass:30 reference-bad items yield 27 rejected and 3 accepted;10 reference-good items yield 8 accepted and 2 rejected.

Reference class	Rejected	Accepted
Bad	27	3
Good	2	8

Your task

1. Calculate overall agreement and each conditional error rate.
2. Explain why overall agreement alone hides the poorer good-item result.
3. State what remains unknown about repeatability.

Use the next page to record your reasoning before reading the answer.

Your practice worksheet

A new matrix with different class sizes

Reference group counts

Correct cells

False-accept calculation

False-reject calculation

Next study question

Complete your reasoning before turning to the answer. Use the separate learner workbook for group practice.

Check your reasoning and transfer it

1. Overall agreement=(27+8)/40=87.5%. False acceptance among bad is 3/30=10%; false rejection among good is 2/10=20%.

2. The unequal reference groups weight the overall result differently. A single pass still cannot establish within-appraiser repeatability; blinded repeated classifications are needed for that question.

Suggested completed record

Metric	Numerator / denominator	Result
Agreement	35/40	87.5%
False acceptance	3/30	10%
False rejection	2/10	20%

Plausible wrong turns

1. 92.5% single-pass agreement is not repeatability.

2. The chosen reference mix is not automatically the production mix.

Evidence of a sound answer

1. Use the correct reference-group denominators.

2. Explain each error direction.

3. Keep reference uncertainty and repeated-study design visible.

Take the learning back to work

Commitment	Agreed follow-through
Owner	Inspection-method owner
Record	Reference basis, item-level classifications, agreement/error summaries and action record
Review	After instruction changes and during planned measurement review
Verification evidence	Representative blind comparisons and resolved decision criteria
If unresolved	Clarify references or method, coach targeted errors and verify with new blinded opportunities.

Help someone else apply the method

Prepare

1. Reference/classification cards
2. Blank matrix
3. Separate answer sheet

Minutes	Activity and coaching prompt
5	Define a matrix cell Which items belong in false acceptance?
8	Compare three percentages What denominator changes?
10	Work unequal groups Which error is hidden by one total?
5	Debrief next evidence How would a repeat be kept blind?

Debrief

1. Ask learners to explain the consequence of each error without inventing a universal threshold.
2. Treat disputed reference examples as method evidence, not automatic appraiser blame.

Individual or remote practice

Color the two matching cells, then calculate each conditional rate from its own reference row.

Connect the method and check its boundaries

Before application

1. Adjudicated reference classifications
2. Representative items and blinded repeats

Outside this lesson

1. Repeatability from a single confusion matrix
2. Universal acceptance criteria

ASQ: Attribute agreement analysis scope

ASQ: Attribute agreement analysis scope

Attribute agreement examines within-appraiser, between-appraiser and reference-standard agreement.

Only the public scope is used; no ISO/TR14468 tables or examples are copied and current standard adoption is not inferred.

Open this lesson and its visual aids