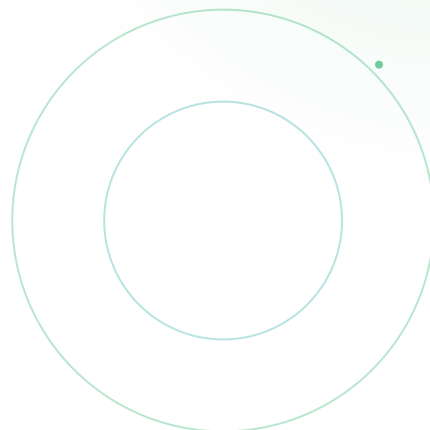




Discrimination Testing Efficiency and the Unspecified Tetrad

Thurstonian power, exact-binomial sample sizes, and why APAC QA labs are moving from triangle to tetrad.

Exact-binomial sample-size tables at working error rates, the reasoning behind the tetrad's one-third assessor saving, and a complete batch-release worked example for dairy and beverage QA programs.

alpha 0.05

One-tailed risk spec

0.80 / 0.90

Power tables

-33%

Typical tetrad vs triangle n

ISO 4120

ASTM E3009-15

ASTM E2262

Exact binomial

BioSensory Pte Ltd

Exclusive APAC distributor — FIZZ Sensory Analysis
(Biosystemès, France)

SENSORY SCIENCE
SERIES
SEPTEMBER 2026

Abstract. Discrimination testing is the workhorse of sensory quality control, ingredient change approval, and packaging migration assessment - yet laboratories routinely run it with more assessors than the question requires, or fewer than the risk justifies. This guide derives sample sizes for the two-alternative forced choice (2-AFC), triangle, and unspecified tetrad tests from first principles using the exact binomial distribution, tabulates them at the error rates working laboratories actually specify ($\alpha = 0.05$; power = 0.80 and 0.90), and explains the Thurstonian reasoning behind the industry's shift to the unspecified tetrad (ASTM E3009-15). A complete worked release-testing example is included.

About this document

This technical whitepaper is published by BioSensory Pte Ltd (Singapore), exclusive distributor of FIZZ Sensory Analysis Software (Biosystèmes, France) across Southeast Asia and Greater China. Internal benchmark figures reflect aggregated, anonymized results from APAC customer deployments and BioSensory pilot studies. Worked examples are illustrative and clearly labeled as such. It is provided for professional education; it does not replace the cited ISO/ASTM standards.

1/3

Guessing probability, triangle and unspecified tetrad

33%

Typical assessor saving of tetrad vs. triangle at matched power

**alpha 0.05 /
power 0.80**

Working-lab default risk specification

1. The Decision the Test Has to Support

A discrimination test answers one question: can a trained-available audience of assessors perceive a difference between two products under controlled conditions? The statistics behind it force a second, commercial question: how much wrongness can we tolerate? A QA release decision that occasionally releases a perceptibly different batch (producer's risk) differs from an ingredient substitution claim that must survive regulatory scrutiny (consumer's risk). Writing down α , β , and the effect size of interest before recruiting a single assessor is the entire discipline of test power - and it is where most laboratories lose efficiency.

2. Tests and Their Guessing Probabilities

Method	Presentation	Question answered	Guessing probability p_0
2-AFC	A vs. B, pick the stronger / target	Directional difference	1/2
Duo-trio	Reference + 2 coded, match the reference	Difference vs. known reference	1/2
3-AFC	3 coded, pick the stronger / target	Directional difference (sharper)	1/3
Triangle	3 coded (AAB), pick the odd sample	Overall difference	1/3
Unspecified tetrad	4 coded (AABB), group the two identical pairs	Overall difference	1/3
Specified tetrad	4 coded (AABB), pick the two A samples	Overall difference (assessor knows A)	1/6

Table 1. Forced-choice discrimination methods and their chance-correct response probabilities (ISO 4120; ASTM E3009-15).

The triangle and the unspecified tetrad share a guessing probability of 1/3: with four samples there are three equally likely ways to split them into two pairs, and exactly one is correct. The tetrad's advantage is therefore not the guessing odds - it is information. Each tetrad observation exposes an assessor to four samples and a grouping decision whose Thurstonian psychometric function rises more steeply with the sensory distance delta than the triangle's. At matched delta and power, published Thurstonian tables report that the unspecified tetrad typically requires roughly one-third fewer assessors than the triangle (Ennis and Jesionka, 2011; Bi, 2015; ASTM E3009-15) - a saving that compounds across a release-testing calendar, because four samples per assessor also fits a session schedule differently than three.

3. Sample Sizes From the Exact Binomial

Tables 2 and 3 give the minimum number of valid (usable) responses n required so that: under no difference ($p = p_0$) the probability of exceeding the critical number of correct responses is at most $\alpha = 0.05$ (one-tailed); and under a real difference expressible as p_1 , the probability of a correct rejection (power) is at least 0.80 or 0.90. Values are computed from the exact binomial distribution, matching the basis of ISO 4120 and ASTM E3009 critical-value tables.

p_1	2-AFC / duo-trio ($p_0 = 1/2$)	Triangle ($p_0 = 1/3$)	Unspecified tetrad ($p_0 = 1/3$)
40%	-	325	325
45%	-	111	111
50%	-	58	58
55%	620	35	35
60%	158	23	23

Table 2. Required number of valid responses, $\alpha = 0.05$ (one-tailed), power = 0.80. Exact binomial computation.

p_1	2-AFC / duo-trio ($p_0 = 1/2$)	Triangle ($p_0 = 1/3$)	Unspecified tetrad ($p_0 = 1/3$)
40%	-	447	447
45%	-	152	152
50%	-	79	79
55%	866	48	48
60%	213	30	30

Table 3. Required number of valid responses, $\alpha = 0.05$ (one-tailed), power = 0.90. Exact binomial computation.

How to read p_1

p_1 is the probability that a single assessor answers correctly when a perceptible difference exists. It encodes the size of the real-world effect: $p_1 = 0.45$ in a triangle test is a subtle difference only strong discriminators catch; $p_1 = 0.60$ is an unmistakable one. Specifying it is a business decision, not a statistics exercise - anchor it to the Thurstonian delta of the product category or to the smallest difference your specification treats as actionable.

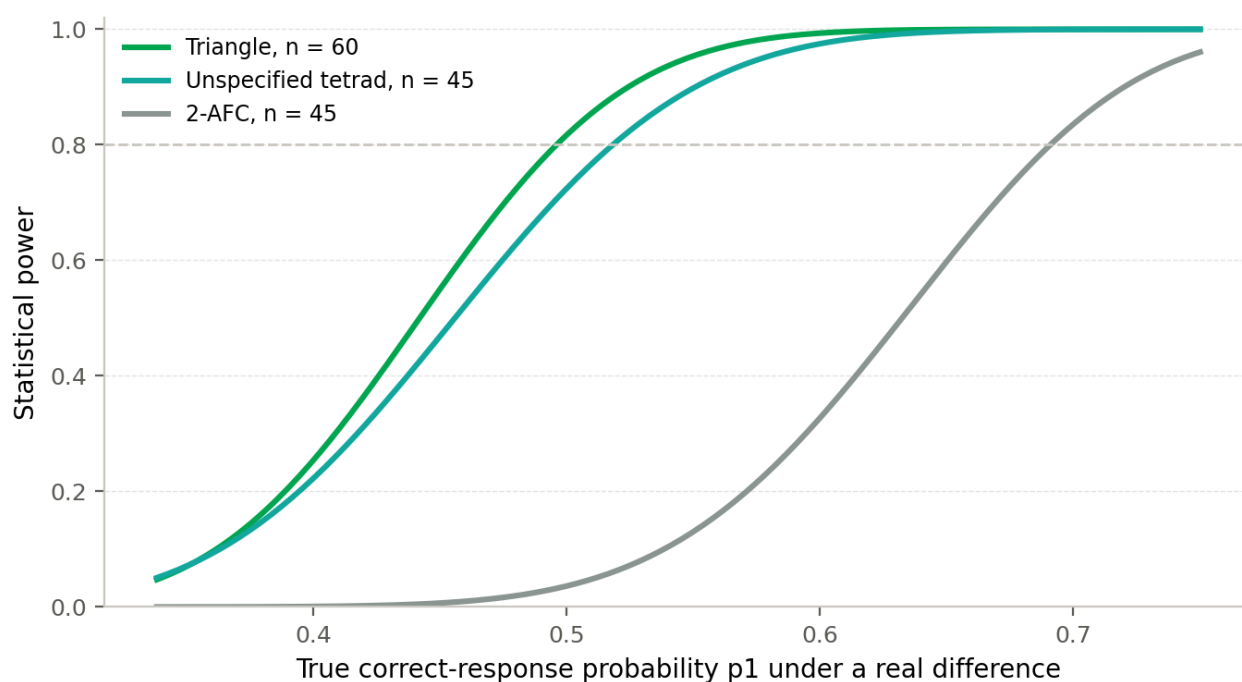


Figure 1. Exact-binomial power curves at $\alpha = 0.05$ (one-tailed). The tetrad reaches target power with fewer assessors across the practically relevant range.

4. Worked Example: Batch-Release Test for a Flavored Milk

A dairy QA team must verify that a packaging change does not create a perceivable difference. The sensory manager specifies: one-tailed test, $\alpha = 0.05$, and wants 80% power to detect a difference strong enough that 45% of assessors would answer correctly ($p_1 = 0.45$). From Table 2: a triangle design requires 111 assessors; an unspecified tetrad requires the same count of valid responses for equal p_1 but captures the difference at higher Thurstonian delta per assessor - and published tables show that when the requirement is written in delta terms, the tetrad delivers matched power with roughly one-third fewer assessors. The team schedules a tetrad with 12 trained internal assessors across 6 replicate panels in FIZZ, tracking completion and invalid responses automatically.

Practical tetrad scheduling note

Four samples per assessor increases sensory fatigue per session. Keep tasting intervals at 30-60 s, provide water and unsalted crackers, cap sessions at 2-3 tetrad sets for foods with lingering aftertaste, and randomize presentation order across all AABB permutations (ASTM E3009-15).

5. Common Failure Modes

- **Recruiting n , then deciding α afterwards.** The critical number of correct responses is fixed by n and α jointly; choosing α after seeing results invalidates the risk claim.
- **Ignoring invalid responses.** Replicates with missed samples must be excluded before analysis; plan recruitment overhead of 10-15% over the tabulated n .
- **Using the triangle where a directional question is meant.** If the brief asks 'is the new process less bitter?', a 2-AFC/3-AFC on that attribute is sharper than an overall-difference test.
- **Treating non-significance as proof of sameness.** A test with 20 assessors and $p_1 = 0.40$ has power far below 0.5; 'no significant difference' may simply mean 'not enough assessors'. Equivalence needs an equivalence design.

References

- [1] American Society for Testing and Materials (2015). ASTM E3009-15 Standard Test Method for Sensory Analysis - Tetrad Test.
- [2] American Society for Testing and Materials (2014). ASTM E2262 Standard Practice for Estimating Thurstonian Discriminal Distances.
- [3] Bi, J. (2015). Sensory Discrimination Tests and Measurements: Sensometrics in Sensory Evaluation, 2nd ed. Wiley-Blackwell.
- [4] Ennis, D.M. (1993). The power of sensory discrimination methods. *Journal of Sensory Studies*, 8(4), 353-370.
- [5] Ennis, J.M., and Jesionka, V. (2011). The power of sensory discrimination methods revisited. *Journal of Sensory Studies*, 26(5), 371-382.
- [6] International Organization for Standardization (2021). ISO 4120 Sensory analysis - Methodology - Triangle test.
- [7] Meilgaard, M.C., Civille, G.V., and Carr, B.T. (2015). Sensory Evaluation Techniques, 5th ed. CRC Press.