The Institute for Perception

For technical reports from The Institute for Perception or for information about short courses, please visit www.ifpress.com or email us at mail@ifpress.com

Reprinted from IFPress (2012) 15(4) 3,4

Proper Task Instructions and the Two-out-of-Five Test

John M. Ennis, Benoît Rousseau, and Daniel M. Ennis

Background: Difference tests continue to play a prominent role as businesses investigate the sensory effects of ingredient or process changes. Even so, the importance of correct instructions in these tests is not always appreciated - incorrect instructions can reduce the ability of a test to detect real differences. The effect of instructions on the outcome of sensory difference testing was recognized when the difference between the 3-AFC and Triangle test was explained[1,2,3] and when the relative power of difference testing methods was explored[4,5]. This topic has resurfaced as other difference testing methods, such as the Two-out-of-Five test, have become more commonly used. In this report, we explain how incorrect instructions can sabotage a method, and we provide guidance for the optimal set of test instructions for the Two-out-of-Five test[6].

Scenario: You work for a supplier of private-label goods to a major grocery chain. In particular, you work extensively on private label beverages, and use unspecified testing methods such as the Two-out-of-Five test to assess potential ingredient changes. In a typical Two-out-of-Five test to compare two products *A* and *B*, respondents are presented with two samples of one product and three samples of the other. The task of the respondent is to correctly categorize the samples and the guessing probability in the Two-out-of-Five test is 1/10. Because you believe that tests with low guessing probabilities are more sensitive, you have been using the Two-out-of-Five test recently as your line of low-calorie beverages has been undergoing reformulation to incorporate a new low-calorie sweetener. In your testing, you have detected very few significant differences between the reformulated and current products. Yet, when you have sent reformulated beverages to your client for approval, your client has rejected several of those reformulations because the existing products were found to be significantly preferred in their own testing program. Your management has asked you to investigate how your own internal testing might be failing to detect these important differences. In particular, you focus on six reformulations. The overall data from these six tests is shown in Table 1.

Flavor	Number Correct	Total Evaluations	Probability of a Correct Response (%)	*p*-value Based on a Guessing Probability of 1/10
Cherry	7	40	17.50%	0.100
Lime	6	36	16.67%	0.145
Orange	7	44	15.91%	0.146
Strawberry	5	38	13.16%	0.330
Raspberry	6	36	16.67%	0.145
Grape	8	41	19.51%	0.048

Table 1. Data from six Two-out-of-Five tests with the instruction "From the five samples presented, identify the same pair".

The Two-out-of-Five Test: There has been a plethora of task instructions listed for the Two-out-of-Five test[7,8]. Four of these instructions include: 1) Select the two samples that are different from the other three; 2) Identify the two samples that are the same as each other and are different from the other three; 3) Select the "same" pair; and 4) Identify the group of two and the group of three. These instructions can lead to different outcomes and have a major influence on the interpretation of results from the test. We will discuss alternatives (3), *Same Pair*, and (4), *Grouping*, to illustrate how instructions critically affect the outcome.

Proper Task Instructions: In order to understand the differences between *Same Pair* and *Grouping*, consider the three cases shown in Figure 1. In each case, we suppose that the two products differ with respect to some variable, such as sweetness, that increases from left to right. When samples of the same product are evaluated, they differ due to random variation in sample perception[9]. Thus we represent the two percepts of the *A* product by *a1* and *a2*, and the three percepts of the *B* product by *b1*, *b2*, and *b3*. In the first case, both *Same Pair* and *Grouping* lead to correct responses. However, in the second case, *Same Pair* gives an incorrect answer even though the samples for the *A* product are both perceived as less sweet than all the samples for the *B* product. In this case, *Grouping* still gives a correct answer. Finally, in the third case, *Same Pair* leads to a correct answer while *Grouping* leads to an incorrect answer. The second case is much more likely to occur than the third, leading to many more correct answers from the *Grouping* instructions than from the *Same Pair* instructions.

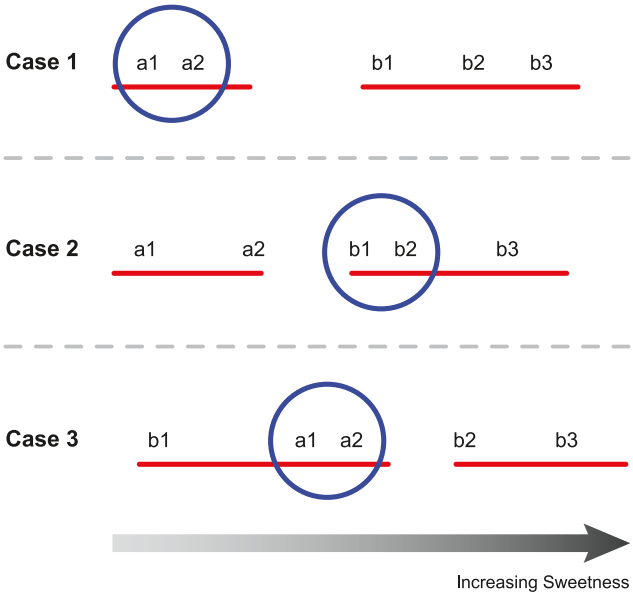

Figure 1. Three cases illustrating the difference between expected results under two instructions - "group the samples into a group of two and a group of three" *(Grouping)* and "choose the same pair" *(Same Pair)*. The *Same Pair* instructions are marked by circling the selected responses and the *Grouping* instruction by underlining the groups.

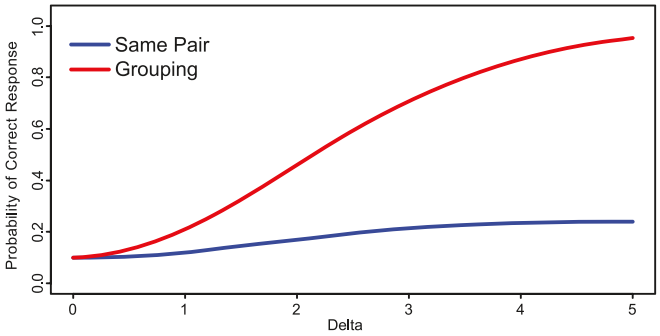

Figure 2. The probability of a correct response as a function of increasing sensory difference. With the *Same Pair* instruction, this probability maximizes at less than 0.25.

Figure 2 shows the probability of a correct response as the size of the sensory difference between the products increases (measured in terms of the standardized measure of difference δ)[9], for both the *Same Pair* and *Grouping* instructions[10]. From this figure, we see that this probability does not approach 100% for the *Same Pair* instructions. In fact, the probability maximizes at less than 25%, handicapping the test against detecting real differences. This fact is reflected in Figure 3, which shows the power curves for the Two-out-of-Five test according to whether the *Same Pair* or *Grouping* instructions are used. Comparison of Figure 3 to similar figures for the Triangle test shows that the Two-out-of-Five test has substantially more power than the Triangle test when the *Grouping* instructions are used, and substantially less power when the *Same Pair* instructions are given.

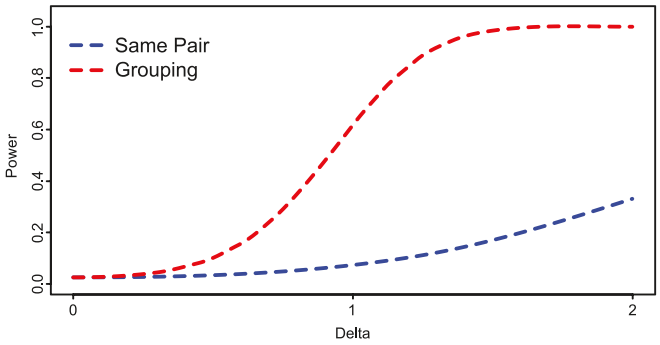

Figure 3. Power curves for the Two-out-of-Five test for the *Same Pair* and *Grouping* instructions. Here $n = 50$.

The Beverage Study with *Grouping* Instructions: In order to explore the theoretical predictions discussed in the above paragraphs using the *Same Pair* instruction, you re-run your experiments under the instruction to classify the samples into a set of two and a set of three. The results are shown in Table 2. Here it can now be seen that the percentages of correct response often exceed 20%, which was not observed in Table 1 and there is evidence of product differences in several cases where one was not detected before. Your concluding decision is to avoid using the *Same Pair* instructions and to adopt the *Grouping* instructions in your future use of the Two-out-of-Five test.

Flavor	Number Correct	Total Evaluations	Probability of a Correct Response (%)	*p*-value Based on a Guessing Probability of 1/10
Cherry	9	41	21.95%	0.018
Lime	8	37	21.62%	0.027
Orange	8	45	17.78%	0.076
Strawberry	7	38	18.42%	0.080
Raspberry	8	38	21.05%	0.032
Grape	9	37	24.32%	0.009

Table 2. Data from six Two-out-of-Five tests with the instruction "Group the five samples into a group of two and a group of three".

Conclusion: Task instructions critically affect the outcome and power of product tests. Sometimes the differences in instructions can produce profound differences in test power. This is certainly the case with the Two-out-of-Five test as there have been many different instructions given for the method, under the apparent misapprehension that all test instructions are equivalent. But, test instructions are important and attention to them leads to a greatly reduced likelihood of missing real sensory differences when using the Two-out-of-Five test. Moreover, the importance of test instructions is not confined to the Two-out-of-Five test - similar findings apply to the Tetrad test as well[11].

References and Notes

1. Frijters, J. (1979). The paradox of discriminatory nondiscriminators resolved. *Chemical Senses,* **4,** 355–358.
2. Gridgeman, N. (1970). A re-examination of the two-stage triangle test for the perception of sensory differences. *Journal of Food Science,* **35,** 87–91.
3. Tedja, S., Nonaka, R., Ennis, D. M., and O'Mahony, M. (1994). Triadic discrimination testing: Refinement of Thurstonian and sequential sensitivity analysis approaches. *Chemical Senses,* **19**(4), 279–301.
4. Ennis, D. M. (1993). The power of sensory discrimination tests. *Journal of Sensory Studies,* **8,** 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. Note that it is common in the sensory literature to refer to methods as "tests", such as the triangle test and the duo-trio test. Since this practice is widespread, we will continue to follow it although it should be understood from the context that we are referring to a method and not a statistical test.
7. Meilgaard, M., Civille, G., and Carr, B. (2007). *Sensory evaluation techniques.* Boca Raton, Florida: Taylor & Francis.
8. Lawless, H. and Heymann, H. (2010). *Sensory evaluation of food: Principles and practices.* New York, NY: Springer.
9. Ennis, D. M. (1998). Thurstonian scaling for difference tests. *IFPress,* **1**(3), 2-3.
10. Ennis, J. M. *(in review).* A Thurstonian analysis of the Two-out-of-Five test. *Submitted to the Journal of Sensory Studies.*
11. Rousseau, B. *(in review).* The importance of correct task instructions in the Tetrad method. *Submitted to the Journal of Sensory Studies.*