Discrimination Testing

A Review of Three Methods:

Maximizing Confidence in Internal Results

Society of Sensory Professionals Conference
October 2010

Janette Pool, Gwen Williams,
Tom Carr, Alexa Williams

FritoLay

Goal of this Research

Examine various tools/methods that can be used for internal discrimination testing

- Compare effectiveness
- Understand the pros/cons
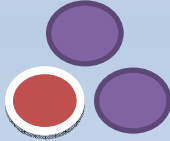- Establish Best Practices and Recommendations

Research Strategy

- 3 Types of Tests
  - Triangle
  - Signal Detection Testing (SDT)
  - Pick-2

- 3 levels of differences
  - no difference
  - moderate difference
  - large difference

- 2 panels
  - Trained
  - Untrained

- 2 product categories
  - Salted Potato Chips (low variability)
  - Seasoned Tortilla Chips (high variability)

34 discrimination tests total!
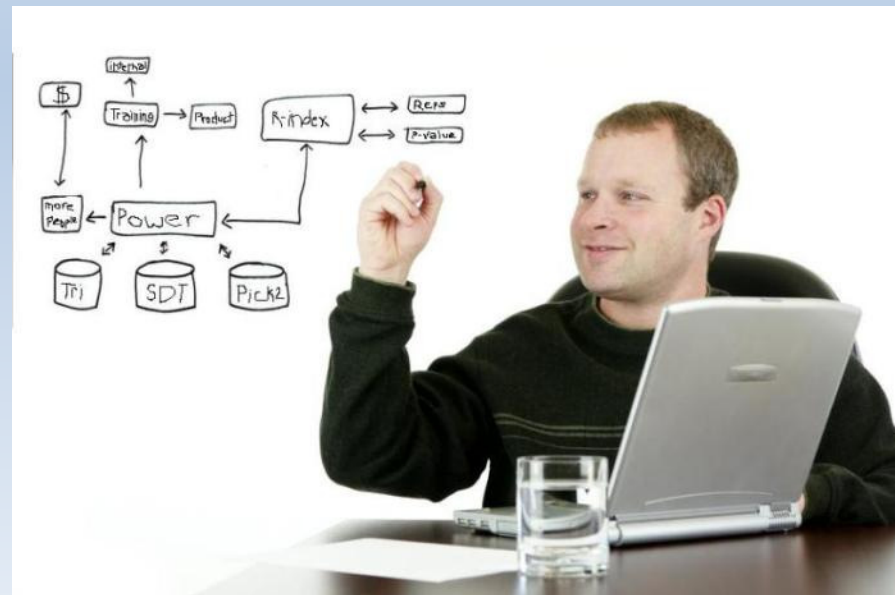
Presentation Flow

- Defining Discrimination Testing
- Overview of each test
  - Triangle Test
  - Signal Detection Test (SDT)
  - Pick-2
- Review other research design details
- Results
- Recommendations

Discrimination Testing – What is it?

Consumer Liking vs. Discrimination

Consumer Liking	Discrimination
Establishing consumer impact of known differences •New and Improved •Equal Liking (Just as Yummy as Ever!) •Competitive Benchmarking	•Are these samples noticeably different?

Frito Lay

When do we use Discrimination Testing?

- Formulation Changes
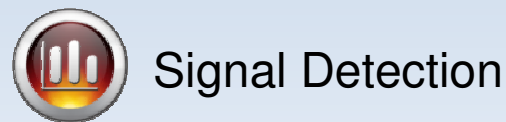
- New ingredient supplier

- Process changes

The ultimate goal is to go unnoticed.
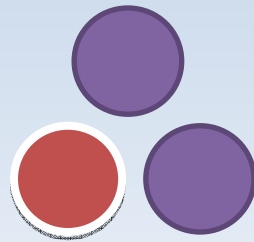

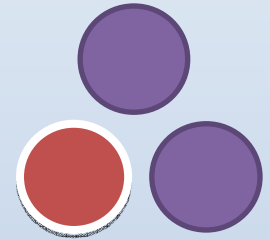Discrimination testing is used to determine if there is a detectable difference between products.

Overview of Methods Evaluated

Triangle

Signal Detection

Pick 2

Triangle Test
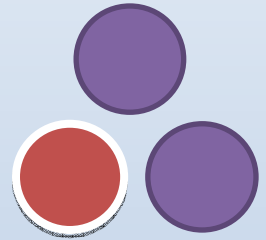
Triangle Test Overview

- Triangle is fairly standard discrimination test method within Sensory Industry.

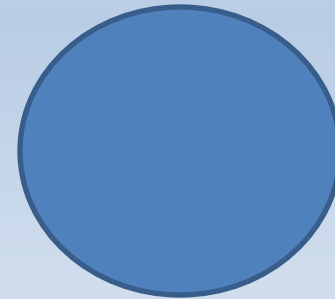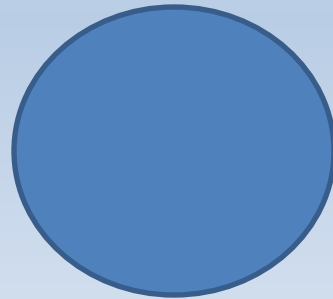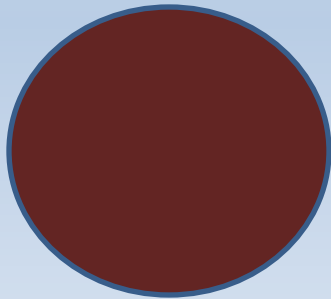"One of these things is not like the other things. One of these things just doesn't belong."
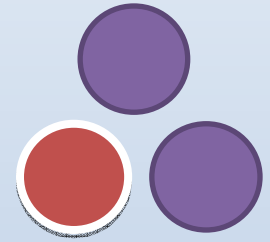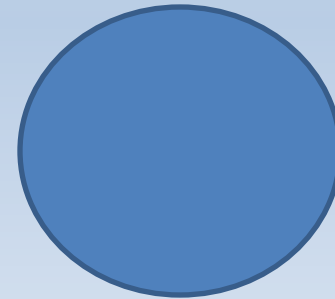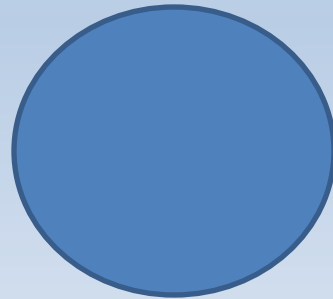
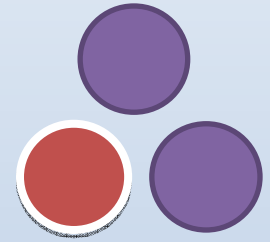# *Triangle Test Overview*

- Evaluator is presented with 3 samples.
  - Two hidden controls
  - One test sample
- Evaluator is asked to select the sample that is different

Triangle Test Example

Triangle Test Example

Analysis for Triangle
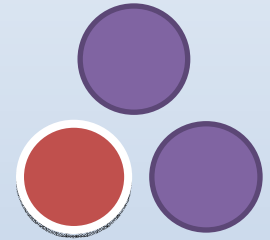
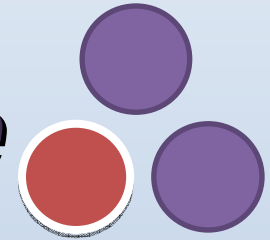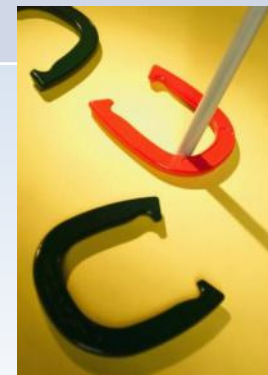- The evaluator has 1/3 chance of getting correct answer by guessing

- The analysis compares the percentage of correct responses vs. expected value of 33%

Pros/Cons of the Triangle

Pros	Cons
• Simple Test • Minimal samples • Widely used	• High probability of guessing (1/3) = limited sensitivity • Ignore product variability

Signal Detection Test (SDT)

Signal Detection Test Overview

- Attempts to eliminate a "response bias" that can result from a forced choice.

 – If forced to make a choice and I'm not really sure, who knows what I will use as the tie breaker.

- Creates a Signal-to-Noise ratio to quantify the magnitude of difference.

SDT: How Does it Work?

- Evaluator is presented with known control

- Test includes several "coded" samples
  - Three hidden controls
  - Test samples (can have 1-6 samples)

- Each sample evaluated sequential monadically

- Evaluator rates how sure he/she is that the sample is Control using 1-4 scale
  - 1: This sample is definitely Control
  - 2: This sample may be Control
  - 3: This sample may not be Control
  - 4: This sample is definitely not Control
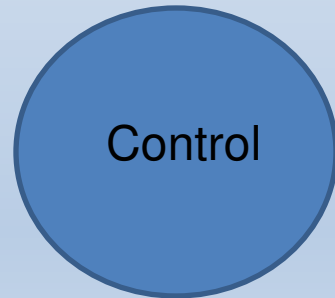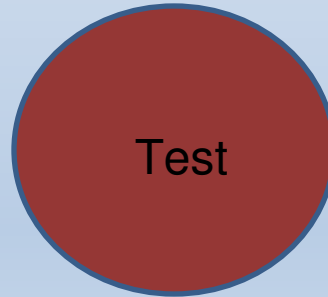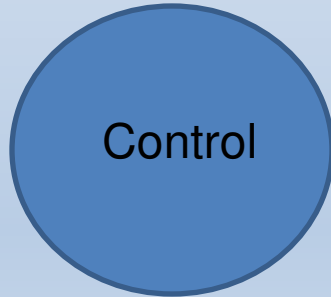
Signal Detection Test Example

Control

Signal Detection Test Example

Control

Test

1: Definitely control
2: May be control
3: May not be control
4: Definitely not control

Signal Detection Test Example

Control

Hidden
Control
#1

1: Definitely control
✔2: May be control
3: May not be control
4: Definitely not control

FritoLay

Signal Detection Test Example

Control

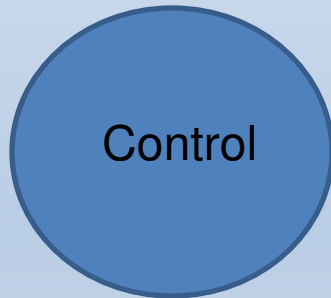Hidden Control #2
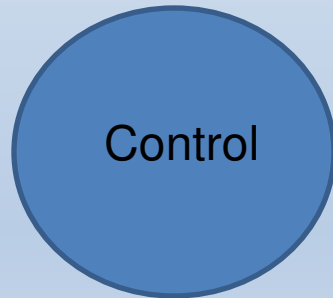
✔ 1: Definitely control
2: May be control
3: May not be control
4: Definitely not control

Signal Detection Test Analysis

Across all evaluators

- Distribution of ratings for **hidden controls** determined ("**Noise**")

- Distribution of ratings for each **test** sample determined ("**Signal**")

- Compare two distributions to create a signal-to-noise ratio called R-index.
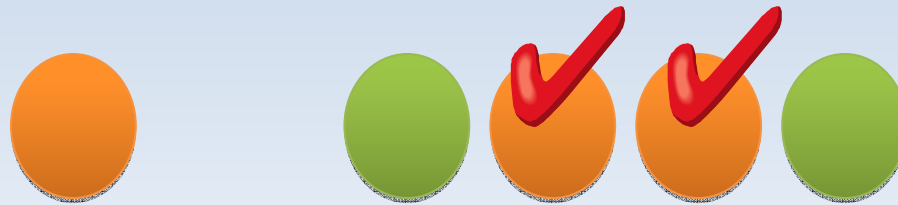
- p-value and d' for the R-index calculated

Pros/Cons of SDT

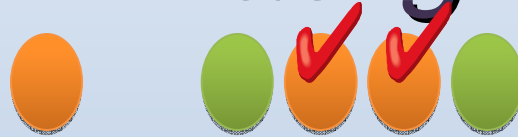| Pros	Cons
•Gives Magnitude of Difference •No guessing or forced choice, "I'm not sure" valid answer •Multiple samples can incorporate product variability	• Test and analysis is more involved and complex • Requires more samples (especially of control)

Pick 2

Pick 2 Background

- Developed internally by Frito Lay in 2005
   - Similar to the method of Tetrads
    - Validated extensively with consumers, n=72

- Existing internal discrimination tests did not always produce results consistent with consumers
    - Large external tests, n=200-300
    - Internal tests said "No Difference"; Consumers said "Different"

- Believed a discrimination test with a lower "guessing rate" would be more sensitive

27
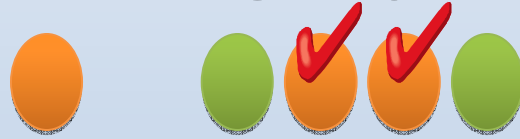
Pick 2 Overview

- Evaluator is given a known control
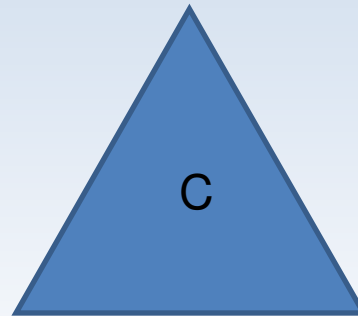
- Evaluator is also given four samples
  - Two hidden controls
  - Two test samples

- The evaluator selects the two samples he/she believes to be closest to the known control.

Pick 2 Overview
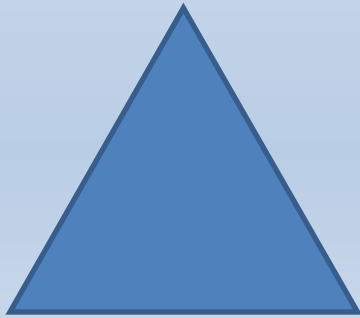
Pick 2 Overview
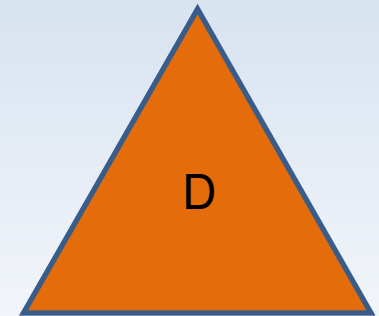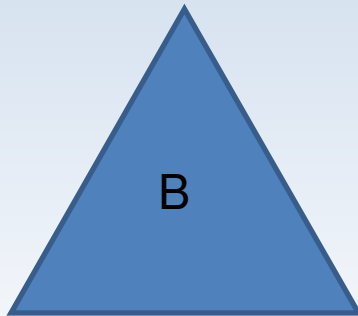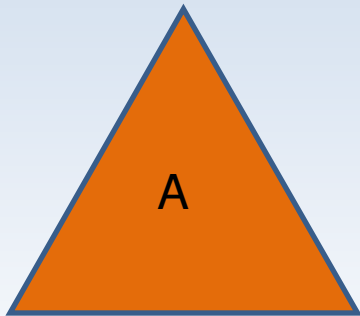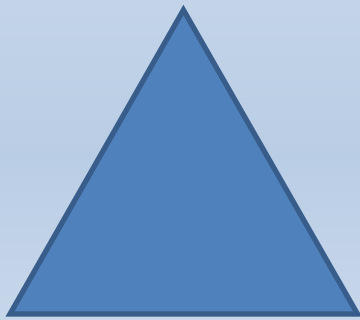
Pick 2 Overview
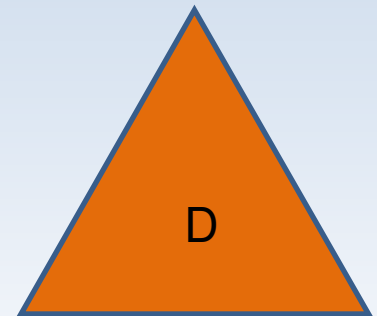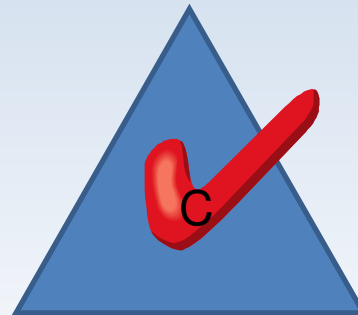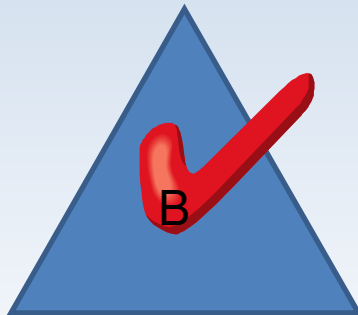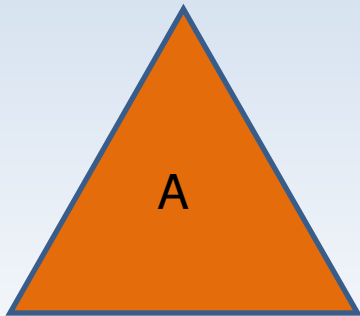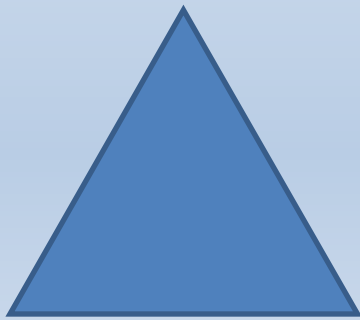
Pick 2 Analysis
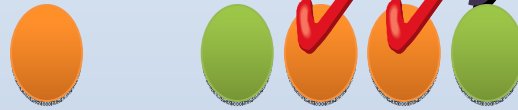
- There is a 1/6 chance of guessing correctly

AB	AC	AD
BC	BD	CD

A   B   C   D
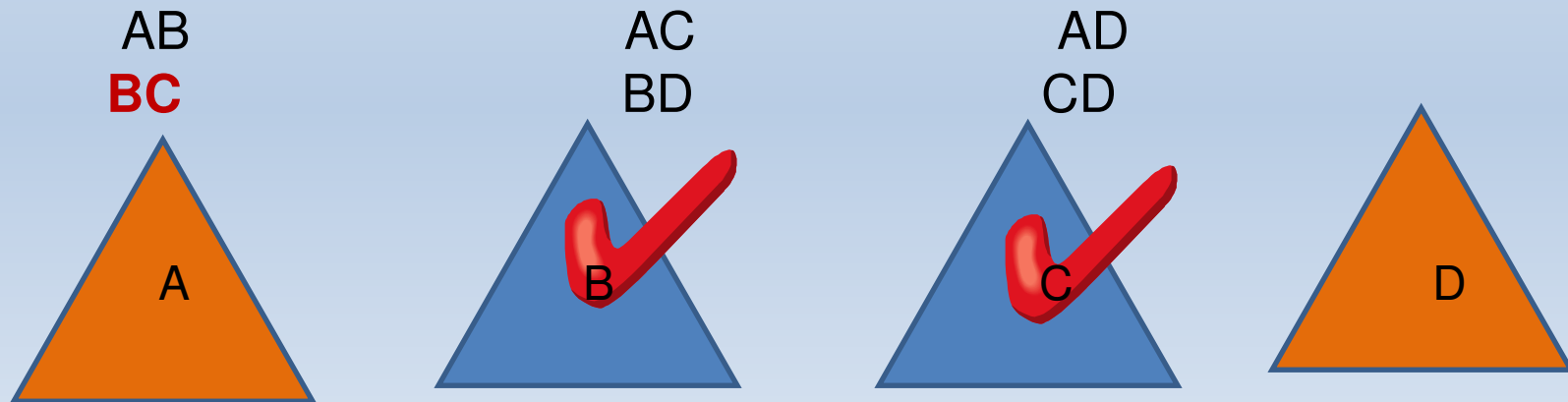
- Analysis compares percentage of correct responses vs. expected value of 1/6 (16.7%)

Pros/Cons of Pick 2

Pros	Cons
• Lower guessing probability so more sensitive • Multiple samples can incorporate product variability	• Test is more complex • Requires more samples
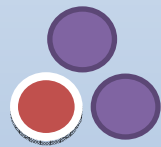
Triangle

Signal Detection

Pick 2

Other Research Design Details
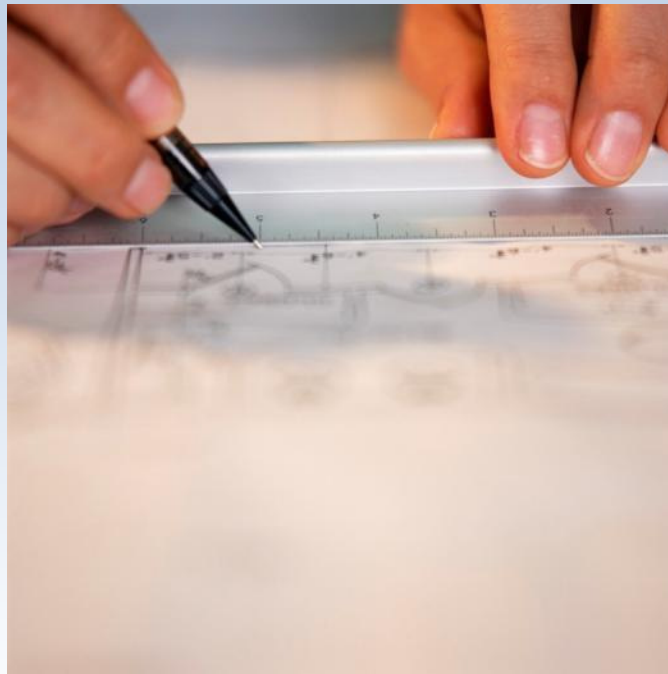
The Evaluators

- Trained Panel (n=10)
  - Trained in Spectrum Method
  - Average 4 yrs experience
  - Same panel used for all tests
  - Had prior experience with SDT, but not Pick 2 or Triangle

- Untrained panel (n=20)
  - Frito Lay employees
  - Screened for product usage
  - Participated in one test per product category

The Products

All testing utilized Salted Potato Chips (PC) and Seasoned Tortilla Chips (TC), both with two levels of toast

How can you be sure about the difference between the products are moderate and large?

Consumer Validation

N=120 frequent users
Each consumer completed two tests: (PC, TC)
Each test contained 3 products: "Control", "Mod", "Big"

Consumer Reaction – Control vs. Moderate Difference

PC

OA: 7.52 / 7.33
FA: 7.46 / 7.25

36 / 64*

TC

OA: 7.61 / 7.55
FA: 7.48 / 7.34

47 / 53
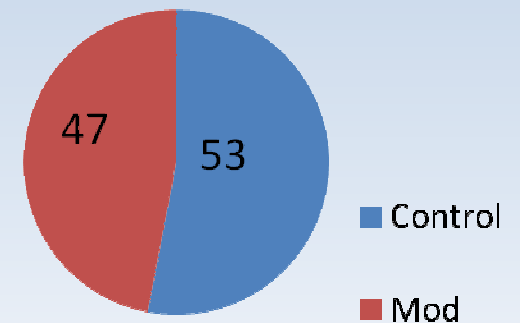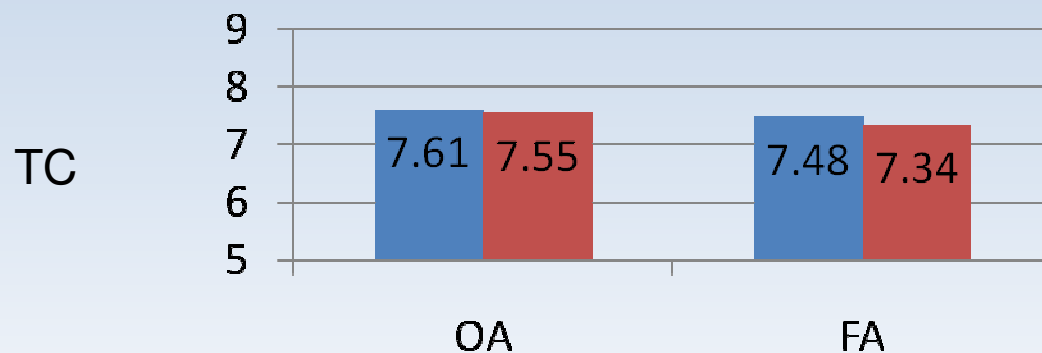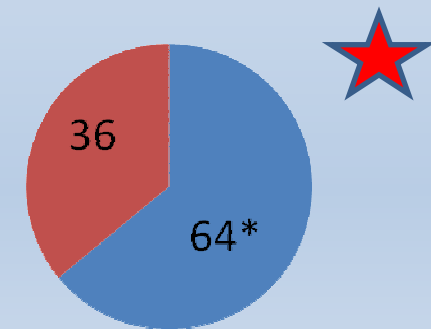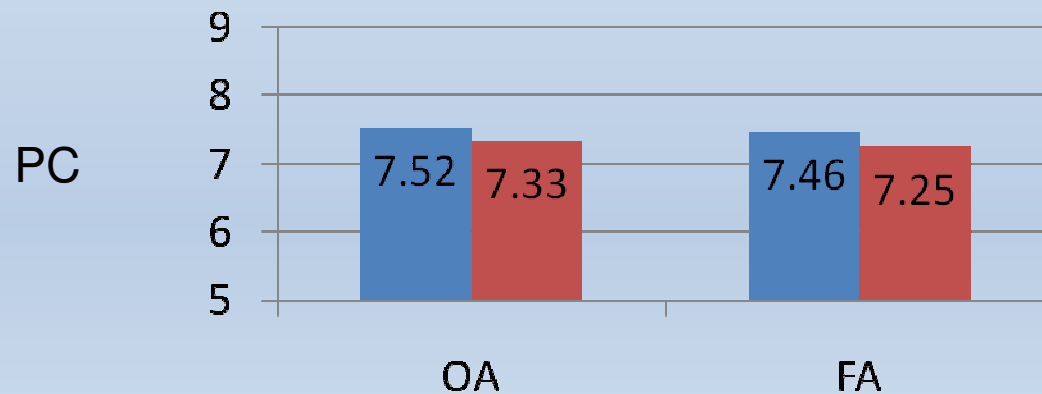
Control
Mod

Preference

For the Product with Moderate Differences:
- Parity liking scores, but directionally lower
- Preference directionally lower, may be significant

40

Consumer Reaction – Control vs. Large Difference
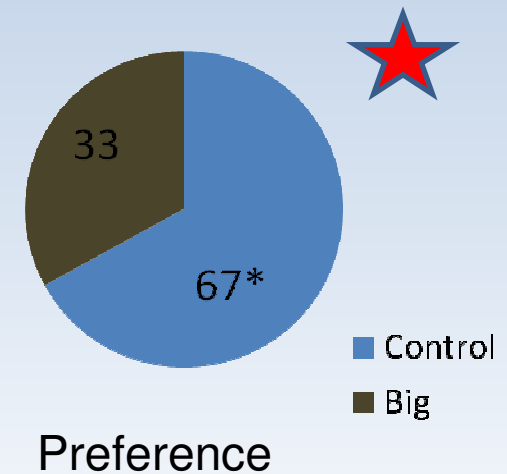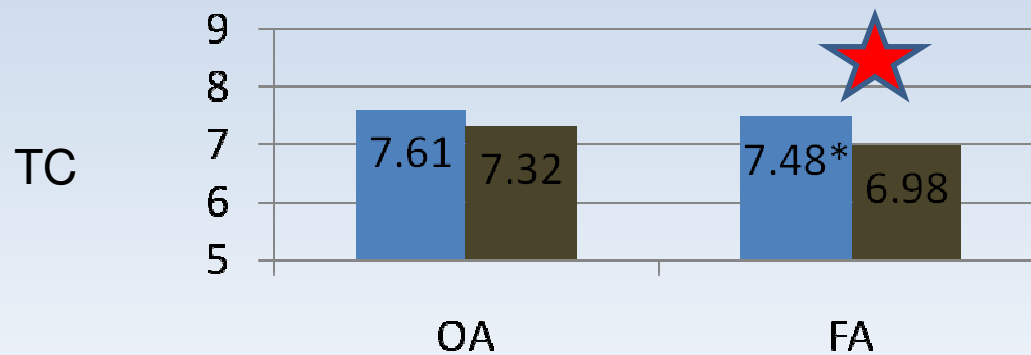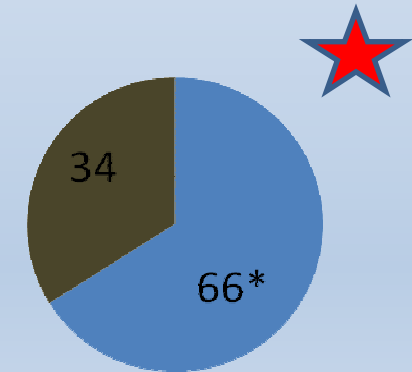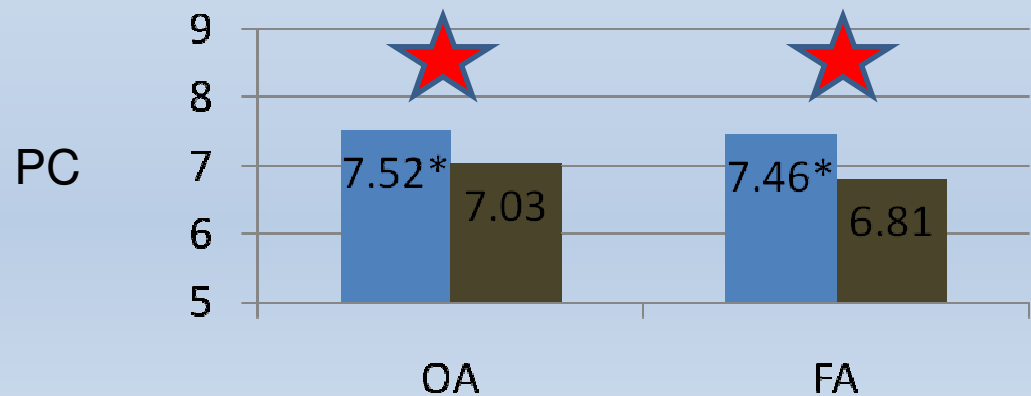
For the Products with Large Differences:
- Significant differences in liking scores
- Significantly lower preference

Consumer Evaluation of Products

Name	Sample Description	Consumer Evaluation
Control	• Control Product. • Representative of in-market design.	
Mod	• Moderate difference from control. • Represents the boundary of acceptable in-market product	• Parity OA, FA • OA, FA, Pref all trend lower
Big	• Large differences from control. • Represents product that would be unacceptable for in-market product.	• OA and/or FA sig. lower • Pref significantly lower

FritoLay

Results

Comparing the Methods – d'

- ## Using d' to compare methods
  - Higher d' values = more sensitive method
  - d' $\geq$ 1 indicates a difference exists
  - d' = infinity notated as d'= 6 for charting purposes

- ## Did we get the correct conclusion?
    - No difference for "control" vs. "control" product
    - Difference for "control" vs. "mod", "control" vs. "big"

Trained Panel
(N=10)

"I'm getting woolly-mammoth notes."

Trained Panel – No Difference
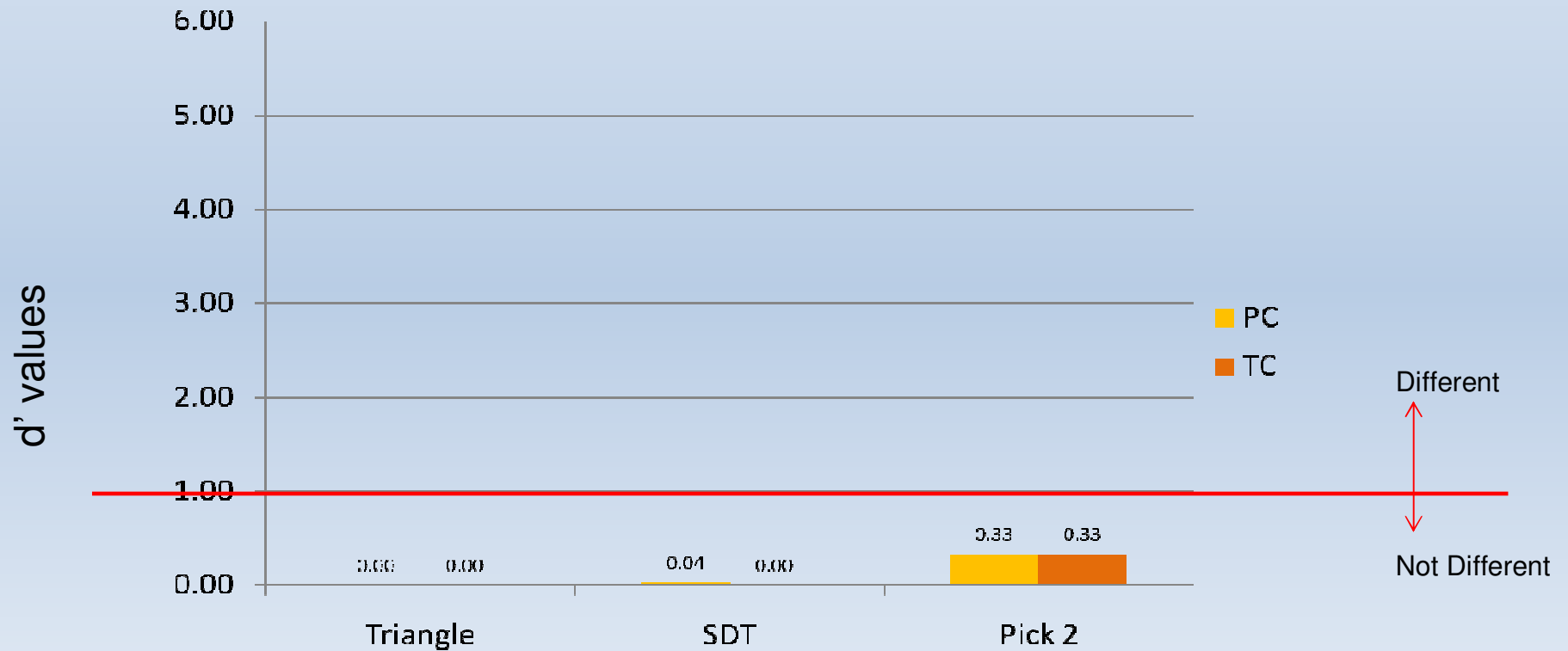
	Triangle	SDT	Pick 2
PC	0.6228 – No Diff	0.4692 – No Diff	0.5155 – No Diff
TC	0.6228 – No Diff	0.6394 – No DIff	0.5155 – No Diff

p-values α = 0.05

- With trained panel, all three tests correctly concluded "No Difference"

Trained Panel – Moderate Difference
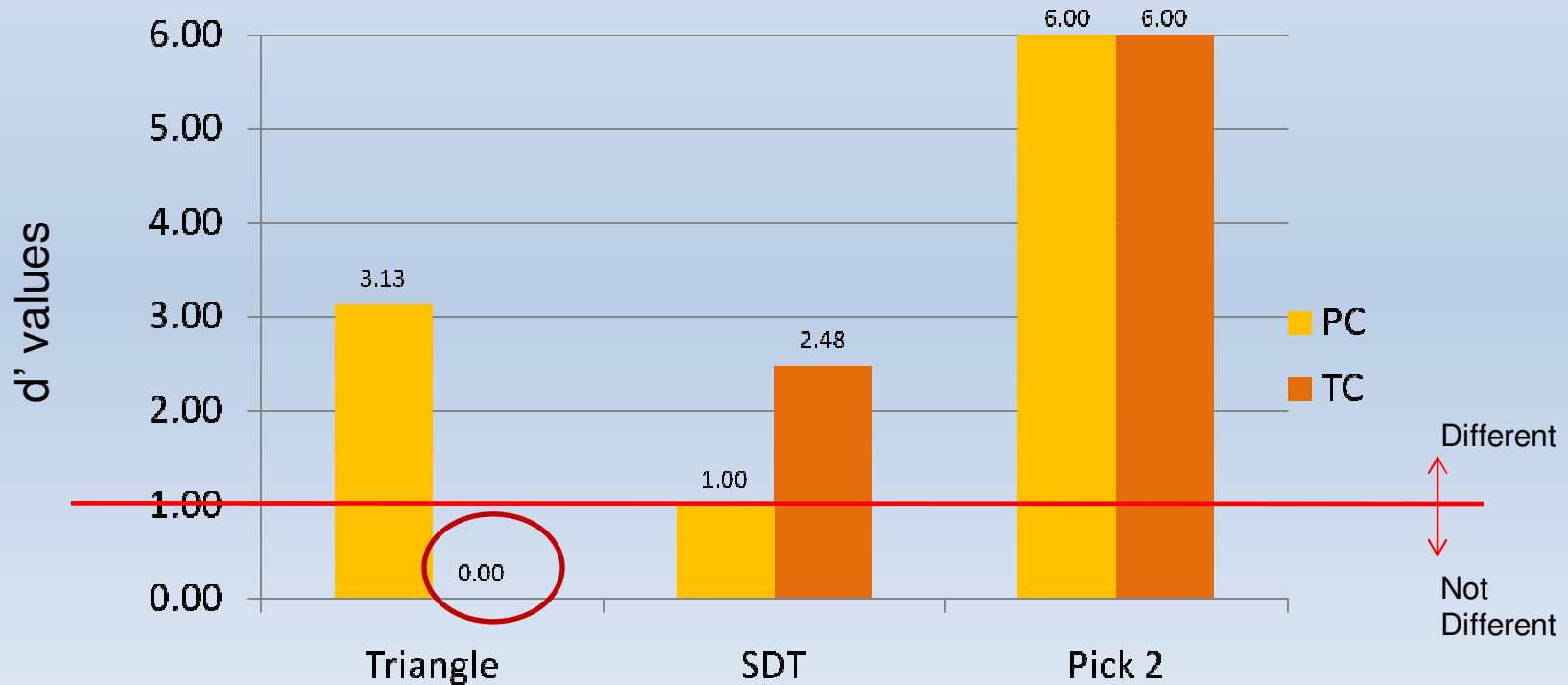
d' values

	Triangle	SDT	Pick 2
PC	0.0034 – Sig Diff	0.0009 – Sig Diff	0.0000 – SigDiff
TC	0.7009 – No Diff	0.0000 – Sig Diff	0.0000 – Sig Diff

p-values α = 0.05

- Pick -2 and SDT had consistently correct results
- Comparing d' values, Pick-2 most sensitive

47

Trained Panel – Large Difference
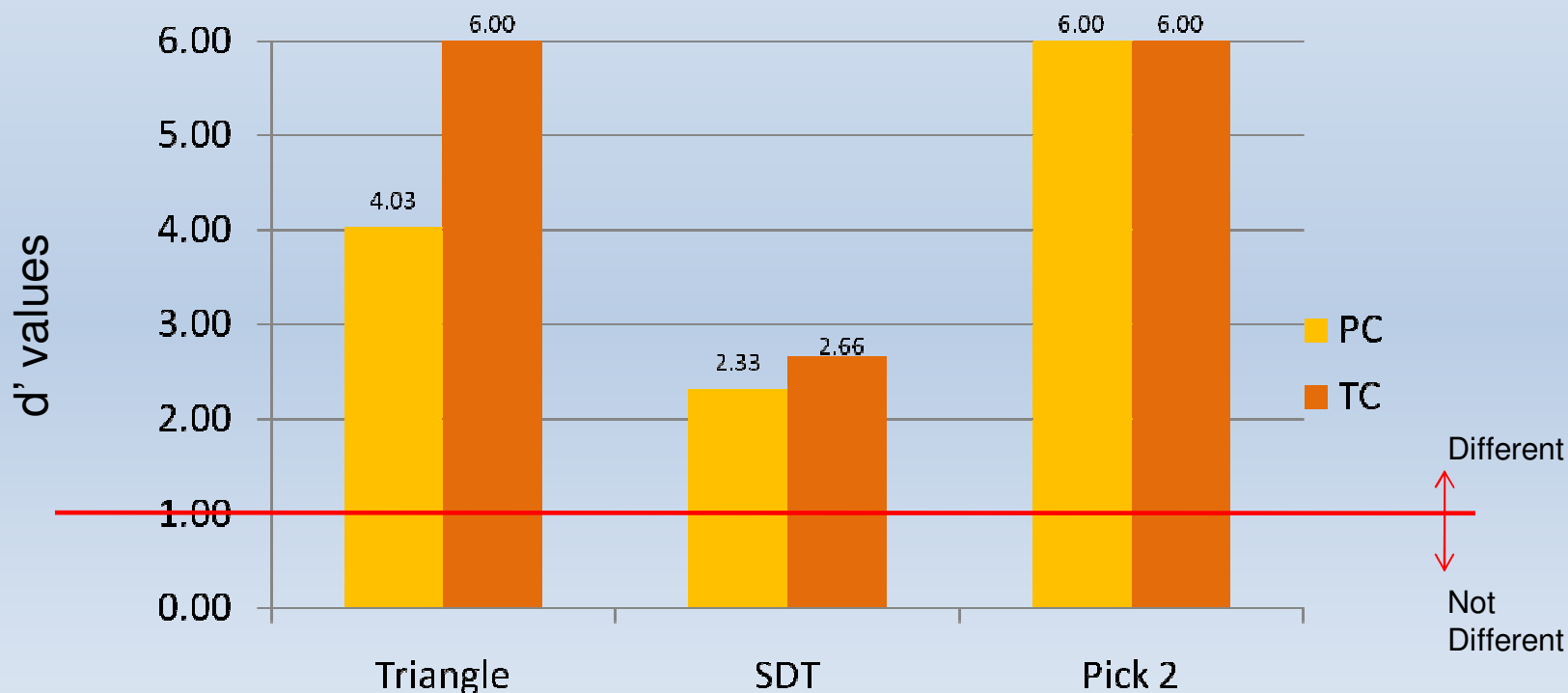
	Triangle	SDT	Pick 2
PC	0.0004 – Sig Diff	0.0000 – Sig Diff	0.0000– Sig Diff
TC	0.0000 – Sig Diff	0.0000– Sig Diff	0.0000– Sig Diff

p-values α = 0.05

- Trained panel detected large difference with ease with all methods
- Pick 2 most sensitive

48

Trained Panel Results

- Pick 2 is most sensitive
- SDT yields correct results and has acceptable sensitivity
- Triangle is not consistently correct (n=10)

Untrained Panel
N=20

Untrained Panel (N=20) – Moderate Difference
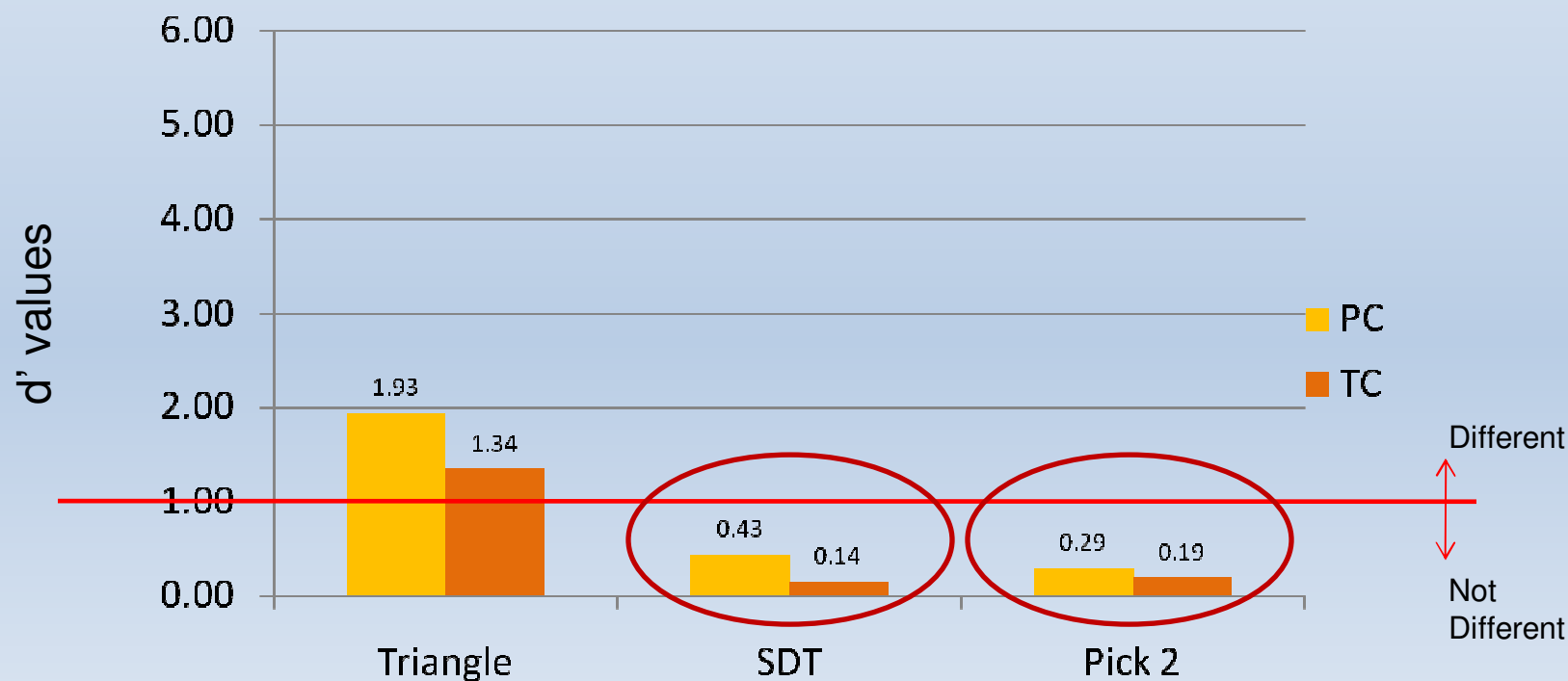
p-values α = 0.05	Triangle	SDT	Pick 2
PC	0.0116 – Sig Diff	0.0140 – Sig Diff	0.2650– No Diff
TC	0.1248 – No Diff	0.2832– No Diff	0.3930– No Diff

- Triangle has highest d' values
  - Anecdotal evidence suggests that SDT and Pick 2 were a more difficult test for untrained evaluators
- p-values suggest that TC results not significant

Untrained Panel – Large Difference
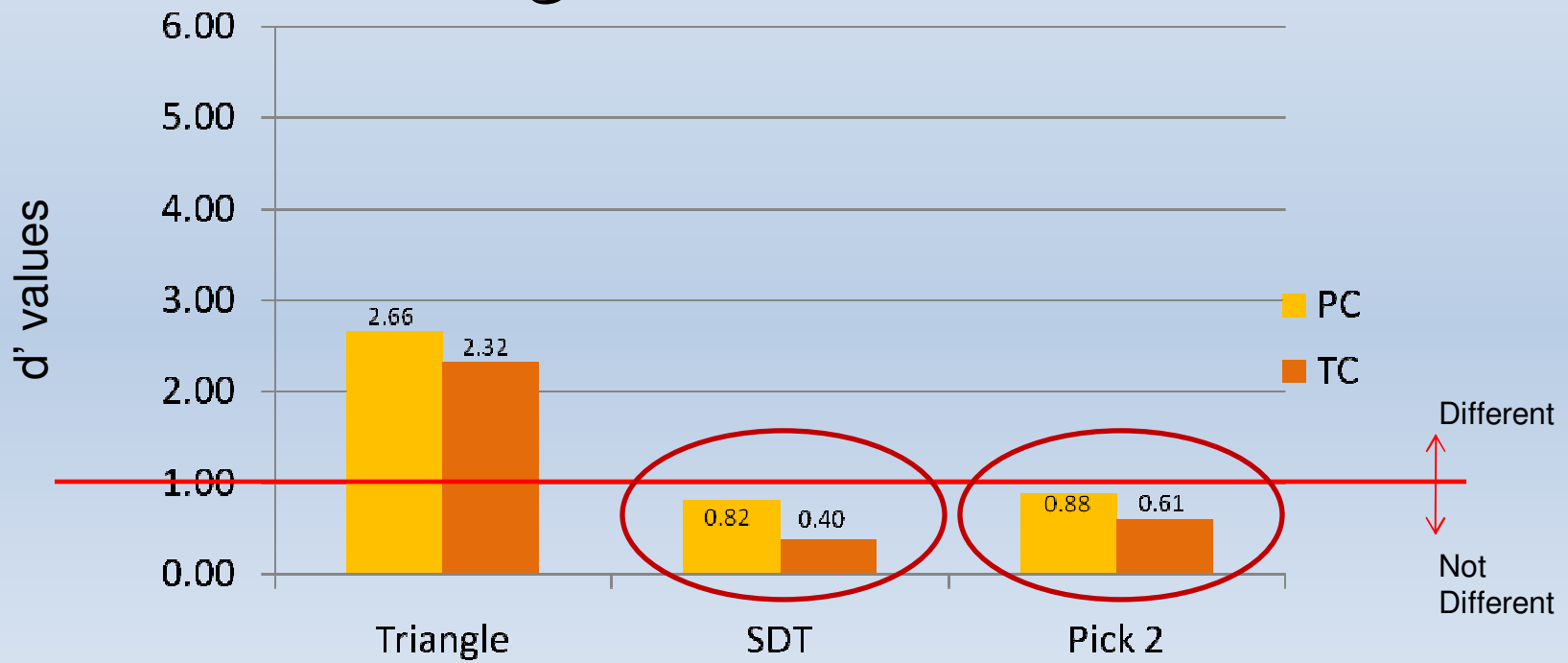
p-values α = 0.05		Triangle	SDT	Pick 2
	PC	0.0002 – Sig Diff	0.0000 – Sig Diff	0.0079 – Sig Diff
	TC	0.0009 – Sig Diff	0.0343 – Sig Diff	0.0479 – Sig Diff

- d' comparison indicates Triangle is only reliable test
- p-value data suggests all results significant

Untrained Panel Results

Solid Line = PC

Dashed Line = TC

🟠 Pick 2

🔴 SDT

🟣 Triangle
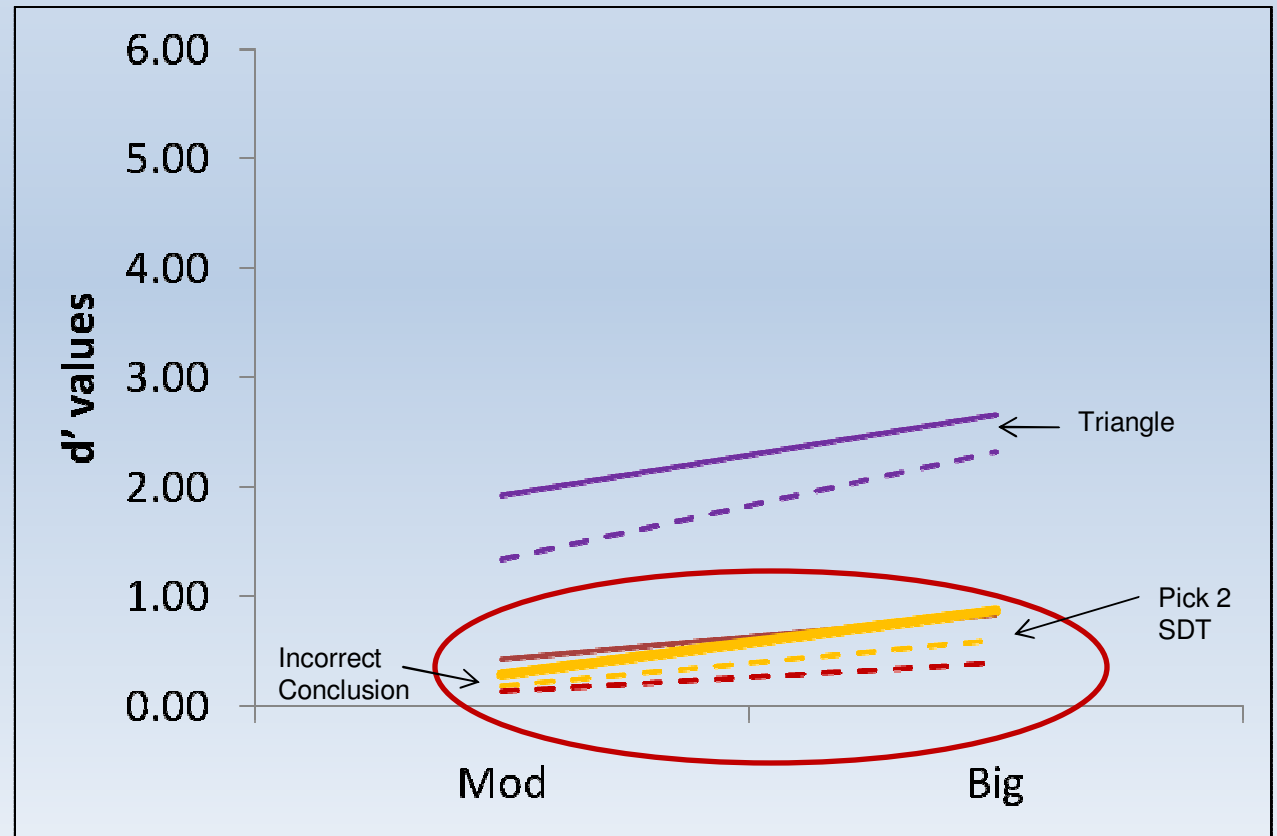
- Triangle test is simple and is the best method for untrained tasters

Trained vs. Untrained Panels

Triangle Test

SDT

Pick 2
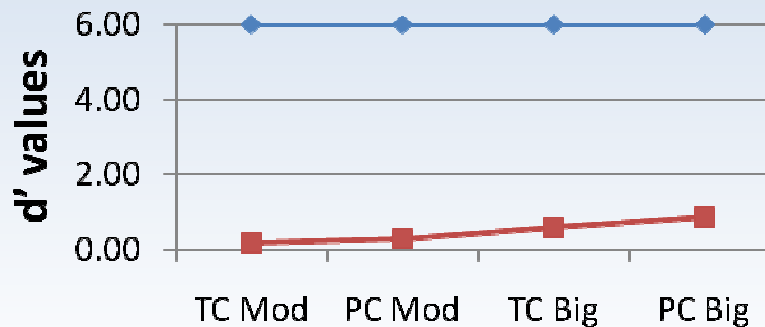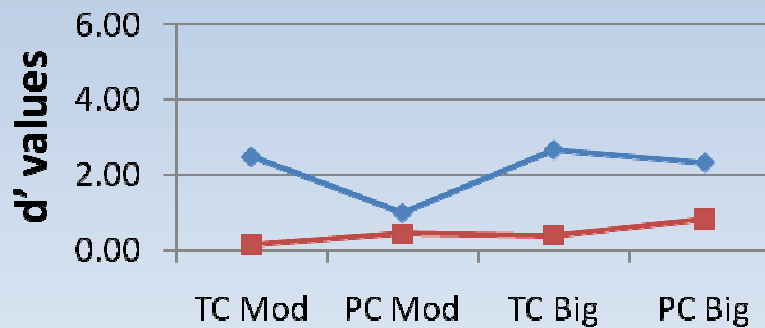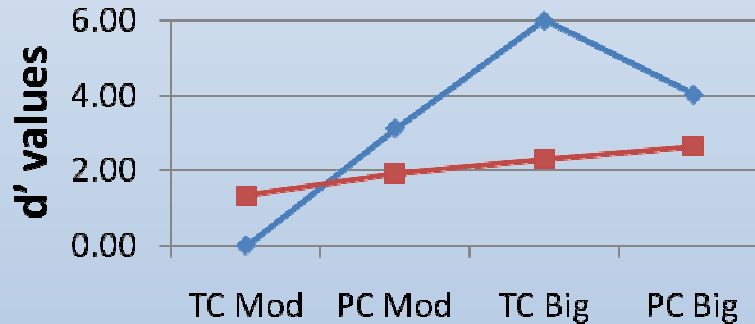
- Trained
- Untrained

- In general, the trained panel is the more sensitive tool for detecting differences, and this is with **_half_** the evaluators of the untrained panel.

- For moderate differences with a highly variable product, a **_low n_**, even with highly trained tasters, is **_risky on a Triangle_**.

SDT – Replicate or Not?

Determining Best Practices for
SDT – Sidebar Research

- The control is replicated three times
- Is there benefit to replicating the test
  samples as well?
  - Replication = more reads = more sensitive
  - Replication = more samples = more fatigue

SDT – Replicate or Not?
(Trained Panel Data)
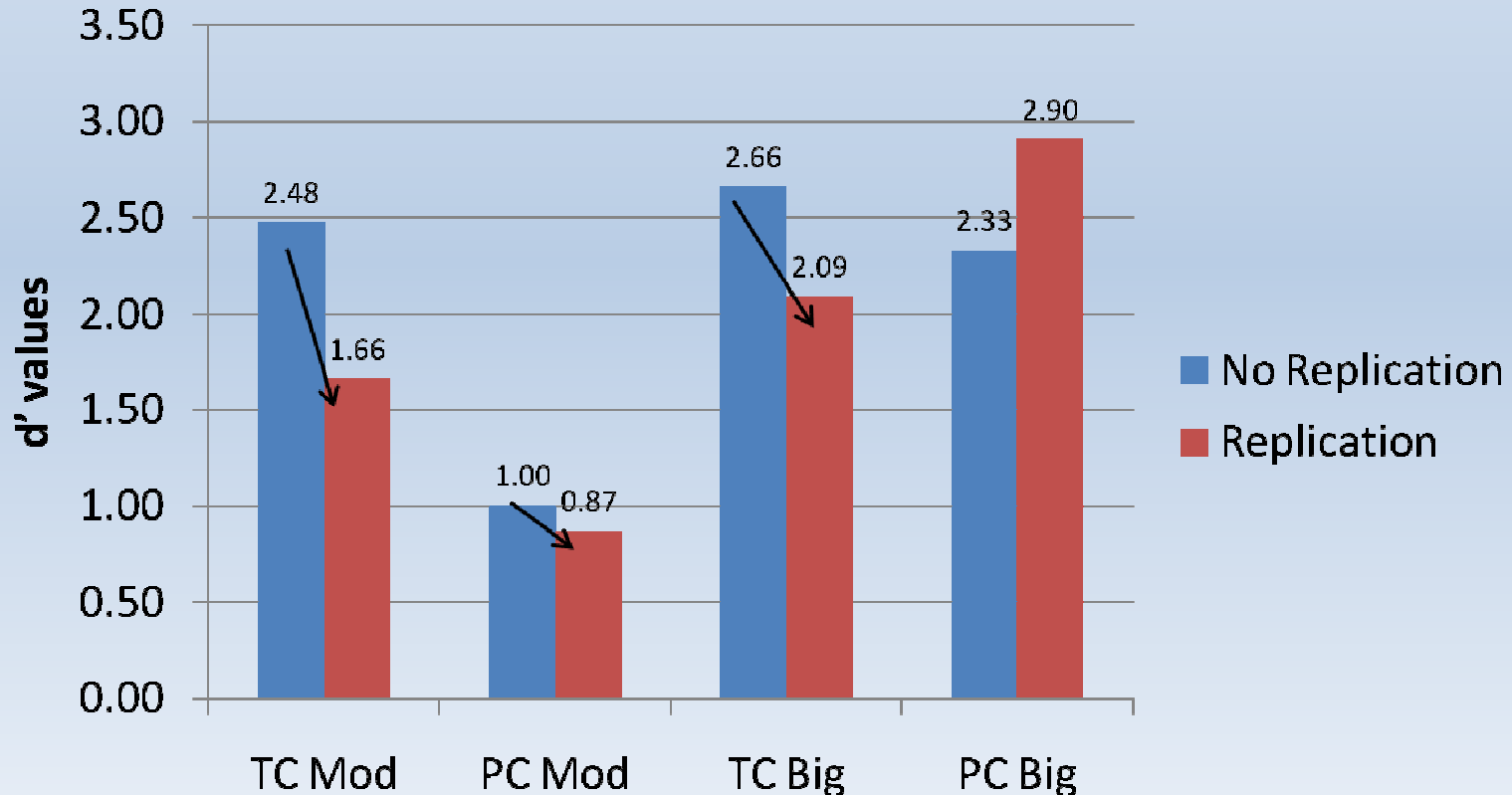
- Replicating the samples does not seem to increase sensitivity, possibly due to fatigue

Recommendations
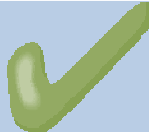
Recommendations

- If trained panel available, use them. They are more sensitive and accurate.
  - Recommend Pick 2 best method for single sample
    - Most sensitive
    - Allows for some product variability to be introduced
  - If have multiple samples to compare, can use SDT
    - Do not replicate test samples

Recommendations, con't

- If Trained panel not available and you must use an untrained panel
  - Use triangle to keep test simple
  - Use more than 20 respondents (minimum of 36 typical rule of thumb)
- In addition to p-values to determine statistical significance, have guidelines to establish meaningful differences
  - % Detectors or d' for Triangle, Pick 2
  - R-Index for SDT

Recommendations – Quick Reference Chart

	Triangle	SDT	Pick-2
	Pros: Simple, only 3 samples Con: May not be as sensitive	Pros: Can do multiple samples; includes variability Con: More complex task and analysis; may see effects of fatigue	Pros: Most sensitive; includes some variability Con: More complex task
Untrained evaluators	✔		
Single sample to compare			✔
Have several pulls or multiple samples		✔	

Frito Lay

Special Thanks

- Co-Authors Alexa Williams, Tom Carr, and Gwen Williams

- Our internal Diagnostic Panel

- Kristine Guidry/Inside Taste – Recruiting our untrained panelists

Questions?

References

Ennis, D.M. (1993). The Power of Sensory Discrimination Methods. *The Journal of Sensory Studies, 8,* 353-370.

Ennis, J.M., D.M. Ennis, D. Yip, & M. O'Mahony. (1998). Thurstonian Models for the Variants of the Method of Tetrads. *British Journal of Mathematical and Statistical Psychology, 51,* 205-215.

O'Mahony, M. (1992). Understanding discrimination tests: a user friendly treatment of response bias, rating and ranking R-index test and their relation ship to signal detection. *Journal of Sensory Studies, 7,* 1-47.

Meilgaard, M., G.V. Civille, & B.T. Carr. (2005). Overall Difference Tests: Does a Sensory Difference Exist Between Samples? In *Sensory Evaluation Technique*, 4th ed., CRC Press, 59-98.

Rousseau, B. (2006). Indices of Sensory Difference: R-Index and d'. *IFPress, 9 (3),* 2-3.