

Practical Performance Improvement Evaluation Without Statistics

Carl Binder

Senior Partner
Binder Riha Associates
www.Binder-Riha.com

Randy Finrock

President and CEO
Incentas, LLC
www.Incentas.com

Binder Riha Associates

Eval w/o Stats ISPI 2002 1

Overview

- Introduction
- Background
- The Purpose(s) of Evaluation
- Practical Measurement Guidelines
- Non-statistical Evaluation Designs
- Q & A

Binder Riha Associates

Eval w/o Stats ISPI 2002 2

ISPI 2002 -- Evaluation Without Statistics

Who Are We?


- Trainers
- Instructional Designers/Developers
- Performance Consultants
- Human Resources Specialists/Mgrs
- Business Managers
- Researchers
- Who else?

Practitioners?

Binder Riha Associates

Eval w/o Stats ISPI 2002 3

Background

- Less than 5% of ISPI article displays and chapters contain results data (Ogden Lindsley's study)
- 1999 ISPI Think Tank Recommendations 
- Practitioners vs. Researchers
- "Show Me The Money" I and II
- "Measurement Counts!" in *Performance Express*
- **GOT RESULTS?** Campaign
- *Practical* evaluation - "Single-case Design"

Binder Riha Associates

Eval w/o Stats ISPI 2002 4

Why Do We Evaluate?

- To **Validate**
- To Hold or Be Held **Accountable**
- To Make **Decisions**

If you have data for making decisions, you've generally got the other two covered.

Brainstorm:

What Kinds of Decisions? What Kinds of Questions?

- **Did it work?** Did the intervention/program actually produce the desired results – behavior, accomplishments (job outputs), business results?
- **How well?** How BIG a difference did it make?
- **How fast?** At what rate/trend did the change take effect? Are there ANY trends or counter-trends?
- **Better than another alternative?** Does this type of intervention/program work better than some other type? Is it more cost-effective?
- **Is it maintaining?** Was our intervention a momentary blip, or did the results last?

What Can Practical Evaluation Achieve?

- *Support* for investment decisions
- *Ability to predict* performance outcomes
- Evaluation results that managers *value and believe*
- Meaningful evaluation as a *routine of doing business*
- New *insights* about what worked and why

Confidence in Results?

- Do we know our intervention **CAUSED** the effect?
- How **BIG** was the effect?
- How much **BOUNCE** or variability?

*Would a scientist, engineer, accountant, or business manager **BELIEVE** our results?*

ISPI 2002 -- Evaluation Without Statistics

What Do We Want to Avoid?

- “Results” that don’t predict performance outcomes
- Big expenditures of time and money
- Unacceptable interference with interventions
- The perception that *“This is academic research.”*
- Complex statistical designs (or complex statisticians!)
- Others?

Binder Riha Associates

Eval w/o Stats ISPI 2002 9

Some Practical Measurement Guidelines

- Measure the **right things**.
 - *Behavior*
 - *Accomplishments (job outputs)*
 - *Business Results*
- Use **standard countable units** whenever possible.
- Include the **time** dimension.
- Recognize that **rating scales are subjective** – useful only when the desired accomplishments are things such as *“People who think...”* or *“Customers who rate us as...”* Then, *count the people, don’t average their ratings.*

Binder Riha Associates

Eval w/o Stats ISPI 2002 10

ISPI 2002 -- Evaluation Without Statistics

The Foundation of Measurement is Counting with Standard Units!

Binder Riha Associates
Eval w/o Stats ISPI 2002 11

We Can Count....

Behavior Influences

- Training
- Job Aids
- Incentives
- Feedback
- Ergonomics
- Tools
- Coaching
- Goal-setting
- Job Design
- Documents
- etc.

Behavior

- Explaining
- Asking
- Deciding
- Writing
- Speaking
- Finding information
- etc.

Job Outputs

- Proposals
- Problems solved
- Satisfied customers
- Buying decisions
- Signed contracts
- Repaired equipment
- etc.

Business Results

- Productivity
- Profits
- Market share
- Revenues
- Product sales
- Cycle time
- ROI
- Customers who say....
- etc.

Binder Riha Associates
Eval w/o Stats ISPI 2002 12

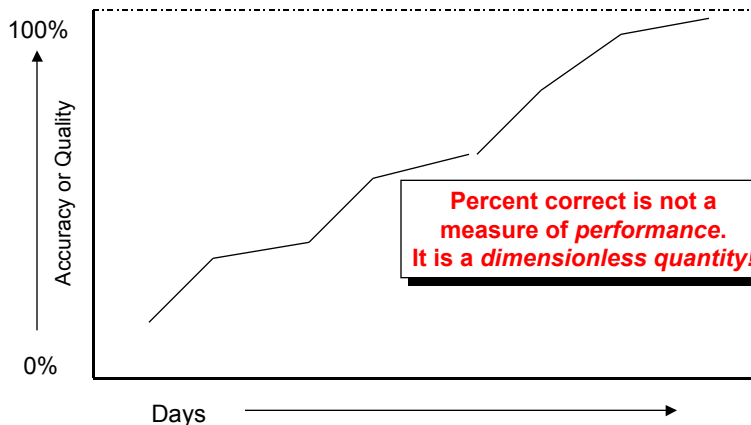
ISPI 2002 -- Evaluation Without Statistics

Countable Units Corresponding to Gilbert's Requirements

Gilbert's Requirements	Corresponding Countable Units
Quality	
Accuracy	Count of accurate items
	Count of inaccurate items
Class	Count of items in each category, rating, or class
Novelty	Count of items defined as novel, in separate categories if desired (e.g., new method, new product, new/better outcome, etc.)
Quantity (or Productivity)	
Rate	Count of any behavior or accomplishment per unit of time (minute, hour, day, week, etc.)
Timeliness	Count of timely events or items
	Count of untimely events or items
Volume	Count of items (as in "sales volume"), or
	Count of volume in units (e.g., liters, cubic yards)
Cost	
Labor cost	Count of dollars spent on labor, by category
Material cost	Count of dollars spent on materials, by category
Management support cost	Count of dollars spent on management support, by category

100% Correct: As Good as It Gets???

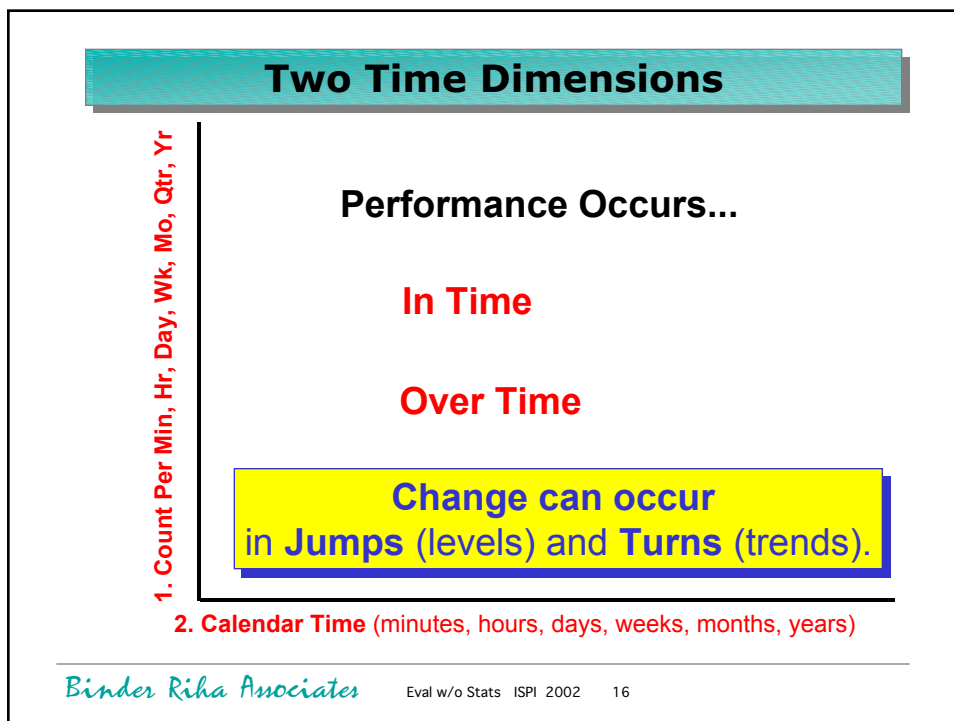
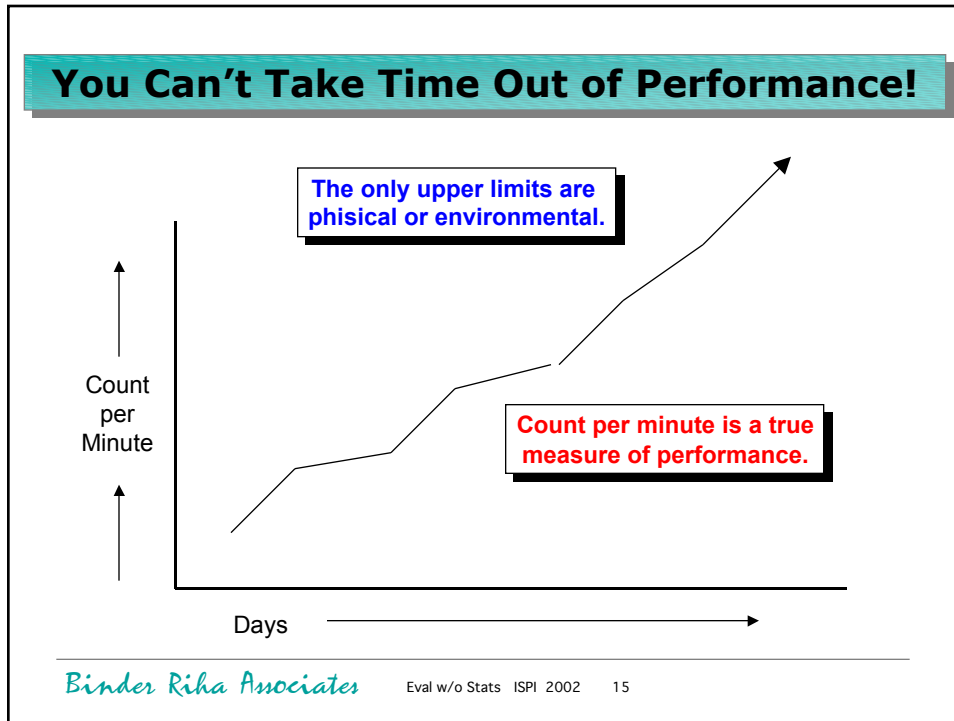
?? "Overlearning" ??



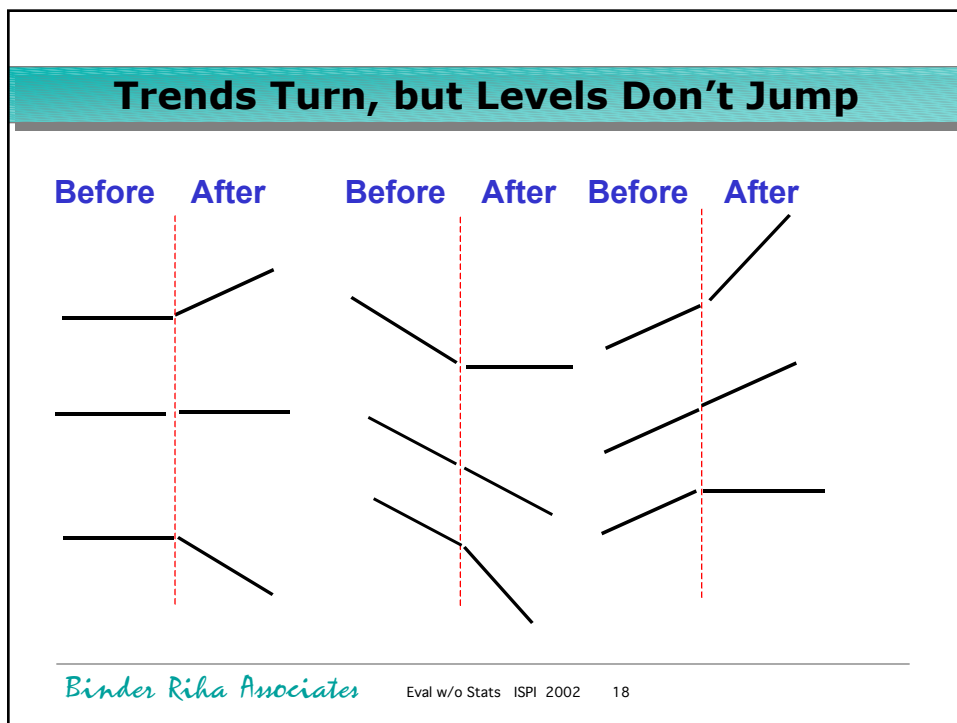
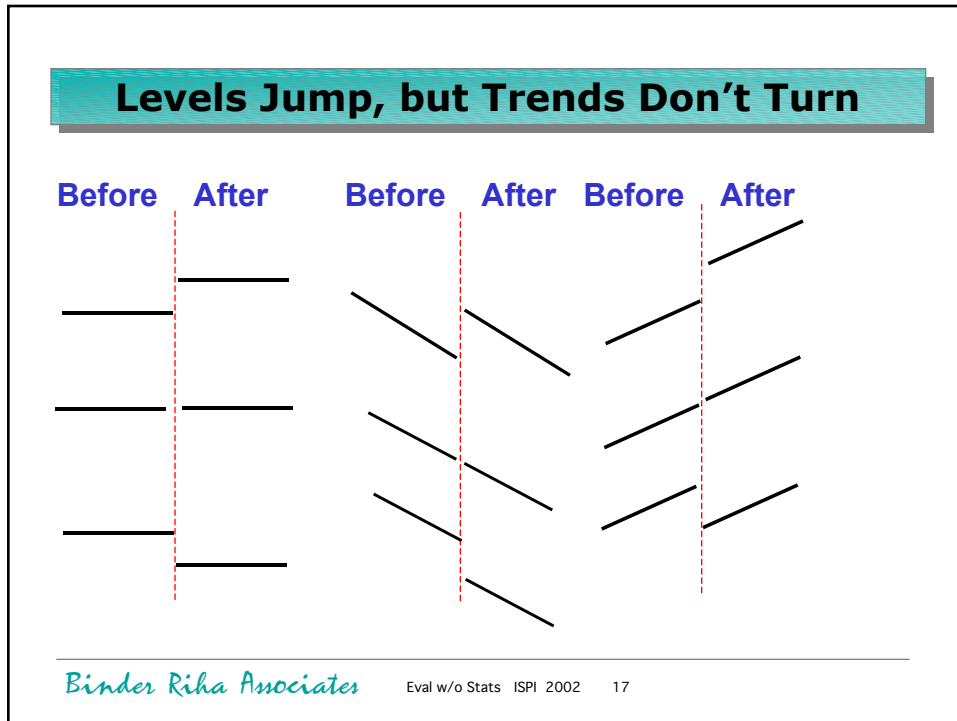
Binder Riha Associates

Eval w/o Stats ISPI 2002 14

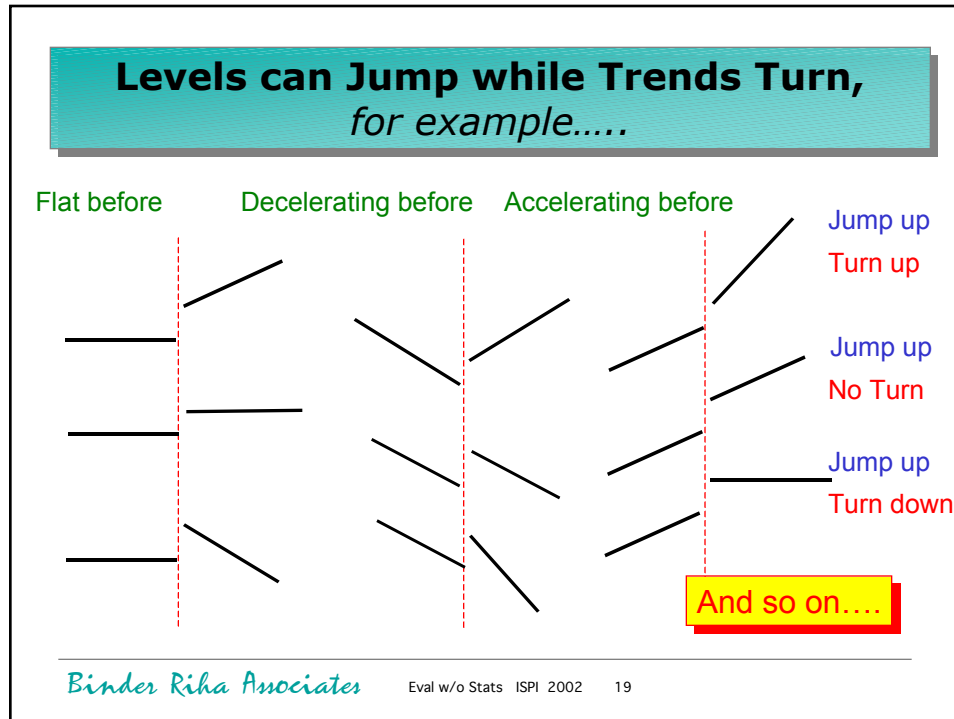
ISPI 2002 -- Evaluation Without Statistics



ISPI 2002 -- Evaluation Without Statistics



ISPI 2002 -- Evaluation Without Statistics



So How Do We Tell What Happened?
Some Simple Evaluation Designs

- **Just measure over time** – “See if it’s changing.”
- **Simultaneous comparison** (groups, settings, etc.)
- **Before -- After** (baseline, change)
- **Reversal** (baseline, intervention, return to baseline)
- **Repeated cases of Before -- After** (replication)
- **Before -- After at different times** (“multiple baseline”)
 - Different individuals or groups
 - Different settings
 - Different measures
 - Behaviors
 - Accomplishments (job outputs)
 - Business Results

Binder Riha Associates Eval w/o Stats ISPI 2002 20

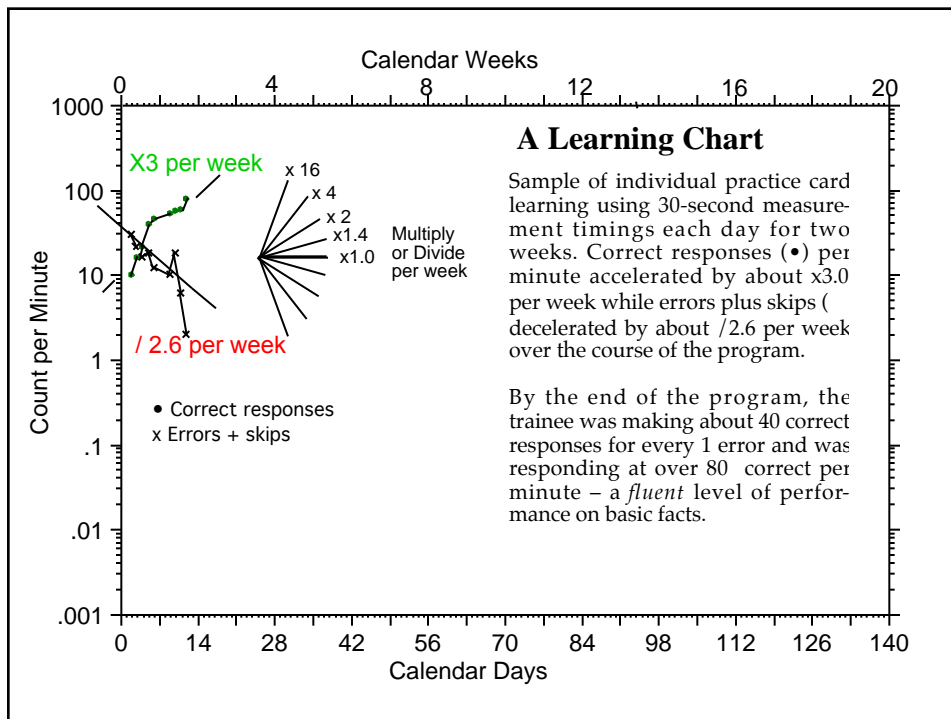
ISPI 2002 -- Evaluation Without Statistics

Just Measure Over Time

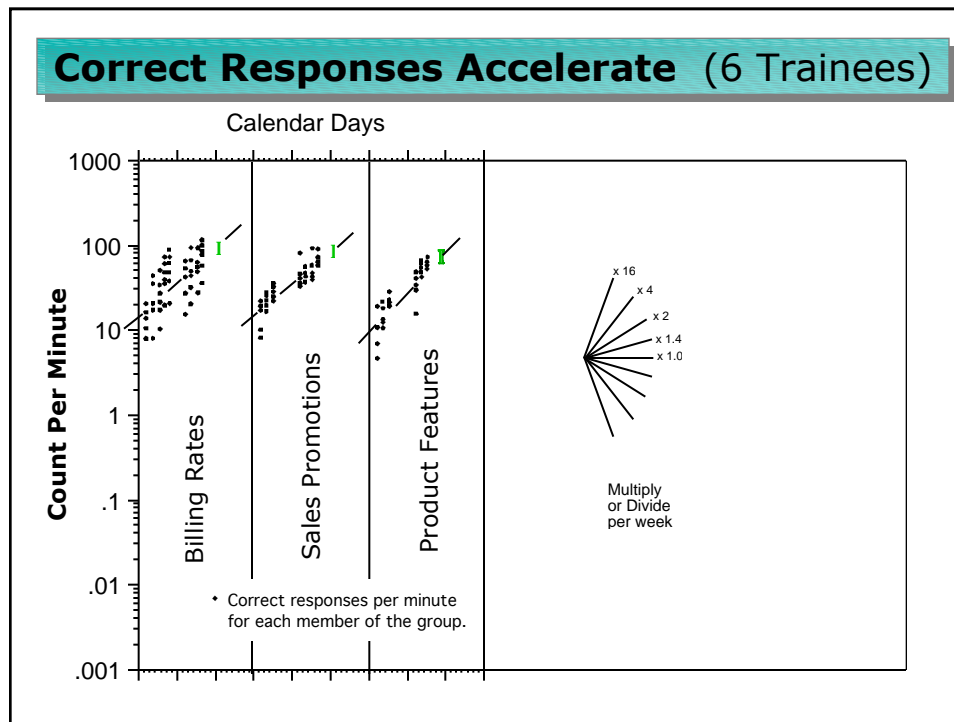
- **Example:**
 - Measuring behavior from the beginning of a program until trainee meets criterion (i.e., measuring learning)
- **When Applicable:**
 - When you don't have a baseline and you just want to see if things are going in the right direction, and if/when they achieve your target
- **Advantages:**
 - Better than nothing
 - If you measure the right thing, you'll be able to see if performance is moving in the right direction
- **Disadvantage:** Can't be certain what caused the change.

Binder Riha Associates

Eval w/o Stats ISPI 2002 21



ISPI 2002 -- Evaluation Without Statistics



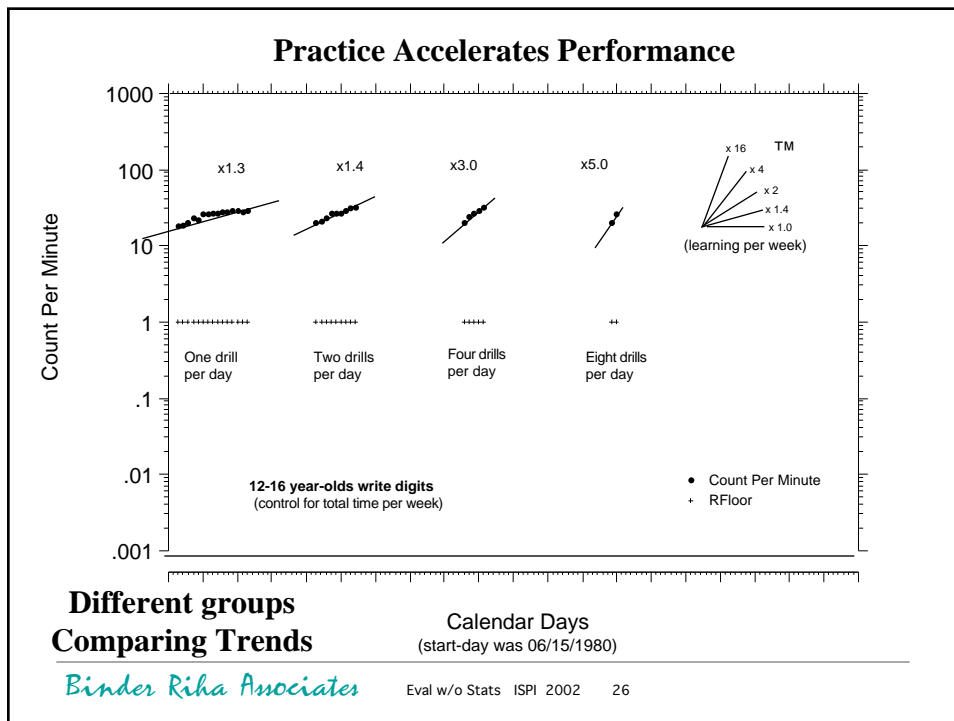
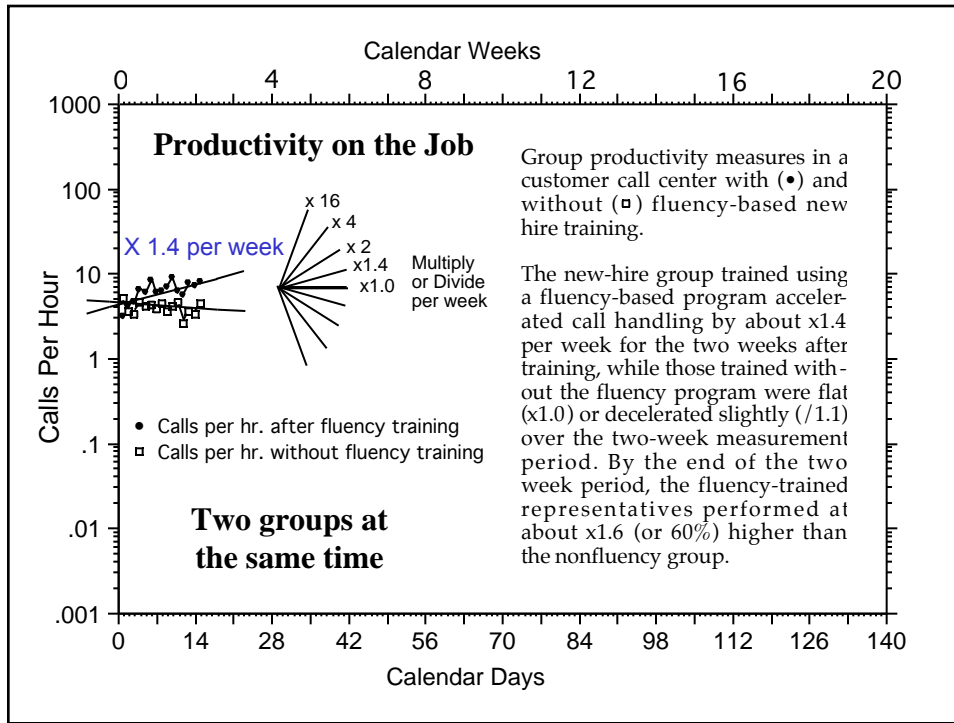
Simultaneous Comparison

- **Example:** Measure productivity over time in multiple groups who receive different performance programs.
- **When Applicable:**
 - When you don't have a baseline
 - If individuals or groups don't have a history
 - When the controlling conditions are easy to identify, and you can be fairly certain that the groups/individuals are comparable.
- **Advantages:**
 - Better than nothing
 - When results are practically significant, this can be the easiest approach and does not require statistics to be convincing.
- **Disadvantage:** Always open to the criticism that the two individuals or groups were NOT comparable.

Binder Riha Associates

Eval w/o Stats ISPI 2002 24

ISPI 2002 -- Evaluation Without Statistics



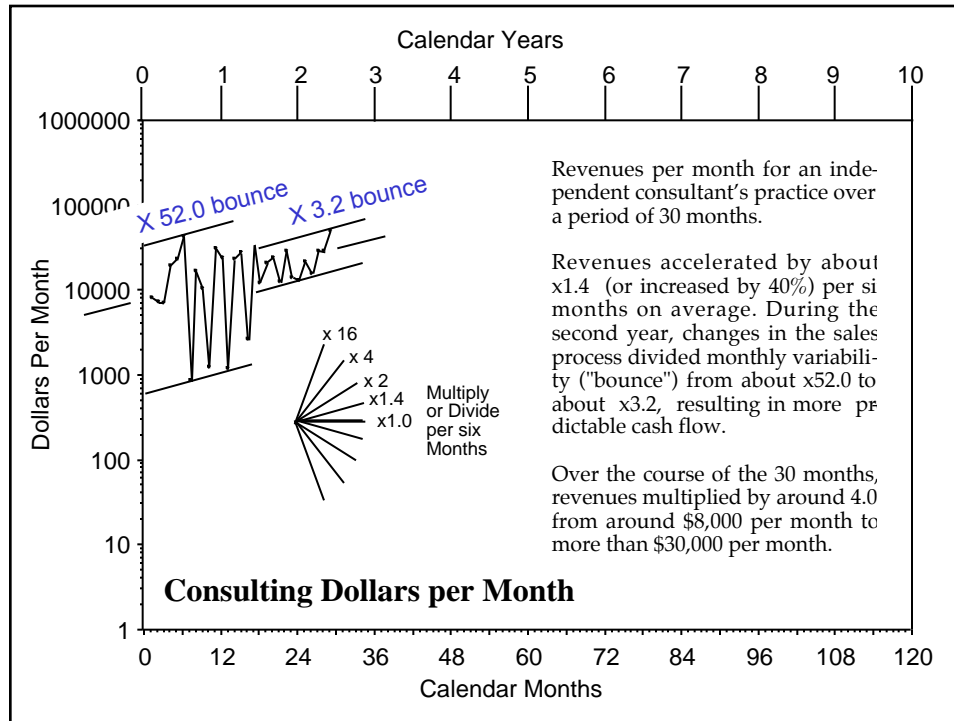
Why Use a Baseline?

- **DEFINITION:** Repeated measures over time for a period BEFORE the intervention.
- **PURPOSE:** To provide a comparison with what happens after the intervention
- **Why REPEATED measures?** Makes us more confident that the effects weren't just *bounce* or an extended trend.

Before – After

- **Examples:**
 - New sales approach compared to previous approach
 - New process compared to old process
- **When Applicable:**
 - When you can obtain baseline measures
 - When the results might not be reversible (e.g., learning)
 - When multiple groups or baselines are not practical
- **Advantages:**
 - Convincing if the results are large
 - Can often use already-existing data
- **Disadvantage:** Open to the criticism that we don't really know what caused the results.

ISPI 2002 -- Evaluation Without Statistics



Reversal to Baseline

- **Example:** Implementing an incentive system, then discontinuing it and continuing to measure.
- **When Applicable:**
 - Seldom applicable in business settings where positive results are desirable and managers would not want to reverse them.
- **Advantages:**
 - If the results return to baseline, it's convincing that your intervention is what made the difference.
- **Disadvantages:**
 - Some effects will not reverse (e.g., learning)
 - Managers will not want to reverse positive outcomes

Repeated Cases of Before – After ("Replication")

- **Example:** Separately measuring sales regions that receive the same intervention at the same time.
- **When Applicable:** When you can gather data for separate individuals or groups, but can't conduct pilot tests or stage sequential roll-outs.
- **Advantage:** Seeing the same effect many times is very convincing.
- **Disadvantage:** "Maybe it was the full moon!" – Open to the criticism that a calendar-related variable accounts for the difference (e.g., "March is always a better month...")

Binder Riha Associates

Eval w/o Stats ISPI 2002 31

Before - After at Different Times ("Multiple Baseline")

- **Examples:**
 - Pilot test with one or more groups before full implementation
 - Roll out programs to different groups spaced over months
 - Use implementation on a series of different "problems"
- **When Applicable:**
 - In any situation where you can implement sequentially across individuals, groups, locations, types of result, or improvement opportunities.
- **Advantages:**
 - Often a very practical fits with pilot-testing or staged rollout
 - When you get results, it is VERY convincing
- **Disadvantage:** Managers may insist on simultaneous implementations without any stages.

Binder Riha Associates

Eval w/o Stats ISPI 2002 32

ISPI 2002 -- Evaluation Without Statistics

Switch to the Overhead Projector!

Binder Riha Associates

Eval w/o Stats ISPI 2002 33

How Might YOU Apply These Ideas?

Binder Riha Associates

Eval w/o Stats ISPI 2002 34

Some Parting Thoughts

- **Statistical significance is not enough.** We need interventions that are PRACTICALLY significant – *large enough to be obvious without statistics.*
- **Multiple baseline designs are practical:** They fit nicely with pilot testing, segmented implementation, and staged program roll-outs.
- **Stretch-to-fill graphs are confusing:** There are advantages to using standard charting methods.
- **We need more sharing:** The more we try these types of designs, and share the results with each other, the more we will learn and improve our practice.

Binder Riha Associates

Eval w/o Stats ISPI 2002 35

References

Publications

- Binder, C. (2001, March) Measurement: A few important ideas. *Performance Improvement*, 40(3), 20-28.
- Daniels, A. (1994). *Bringing Out the Best in People*. New York: McGraw-Hill. Several parts of this book, especially chapter 13, provide simple overviews of basic measurement and evaluation designs.
- Kazdin, Alan E. (1982) *Single-Case Research Designs: Methods for Clinical and Applied Settings*. Oxford University Press. A thorough, somewhat academic, treatment.
- Lindsley, O.R. (1997) Performance is easy to monitor and hard to measure. In R. Kaufman, S. Thiagarajan, & P. MacGillis (Eds.), *The guidebook for performance improvement: Working with individuals and organizations*. San Francisco: Pfeiffer, Jossey-Bass, 519-559.

Links

- www.celeration.org – Web site of the Standard Celeration Society, a resource for use of Standard Celeration Charts and additional references about charting. Also, a charting list server (mostly academics and educators, with a few business people).
- www-personal.umich.edu/~hinderer/scrdrefs.html – A rather unusual resource created by a physical therapist. Contains an extensive reference list on use of “single case designs,” the form of non-statistical evaluation designs upon which this presentation was based.
- www.binder-riha.com – Binder Riha Associates’ site has “Helpful Resources” page.

Binder Riha Associates

Eval w/o Stats ISPI 2002 36