The Wraparound Team Observation Measure: Psychometrics, Reliability, and a Profile of National Practice

Ericka Weathers, MA and Spencer Hensley, BA
University of Washington School of Medicine, Seattle, WA

Jesse Suter, PhD
Center on Disability & Community Inclusion, University of Vermont

Eric Bruns, PhD
University of Washington School of Medicine, Seattle, WA

This project was supported by the Child, Adolescent, and Family Branch of the Center for Mental Health Services, SAMHSA
by the National Institute for Mental Health (R41MH077356; R34MH072759)
And by the Center on Disability & Community Inclusion at the University of Vermont

Goal of Presentation

- Summarize four studies that examine:
  - wraparound practice nationally as assessed by the TOM and
  - the reliability and validity of the measure.
    - Internal consistency
    - Inter-rater reliability
    - Construct validity

Importance of Fidelity Monitoring

- Fidelity is the extent to which a program or intervention adheres to a specified program model.
- Reliably and validly measuring adherence to fidelity is fundamental in ensuring the dissemination and implementation of effective treatments and services.
  (Schoenwald, 2011)

Key aspects of the wraparound practice model, and measurement approaches

- Practice model
  - phases and activities
- Principles
  - cut across activities of the practice model

- Interviews with staff and families
- Team Observation
- Document review

- Organizational and System-level supports
  - without which adherence to the principles and practice model is unlikely

- Key stakeholder survey/interview

Team Observation Measure (TOM)

- The TOM was designed in 2006 to assess adherence to standards of high-quality wraparound during wraparound team meetings.
- It is organized according to the 10 principles of wraparound, with two items dedicated to each wraparound principle.
- Each of the 20 items has 3-4 indicators (71 total), which must be scored:
  - Yes (This was observed)
  - No (This was not observed)
  - N/A (This is not applicable)
- Use of the TOM is supported by a training toolkit that includes a self-test of knowledge of scoring rules and training to criteria using an online video
- The TOM is also supported by an online data entry, scoring, and reporting system (WrapTrack; see www.wrapinfo.org)

Examples of TOM Items and Indicators

Item 4: Effective Decision Making
a. Team members demonstrate consistent willingness to compromise or explore further options when there is disagreement.
b. Team members reached shared agreement after having solicited information from several members or having generated several ideas.
c. The plan of care is agreed upon by all present at the meeting.
d. The facilitator summarizes the content of the meeting at the end of the meeting, including next steps and responsibility.

Item 17: Focus on Strengths
a. Team members acknowledge or list caregiver/youth strengths.
b. Team builds an understanding of how youth strengths contribute to the success of team mission or goals.
c. In designing strategies, team members consider and build on strengths of the youth and family.
d. Facilitator and team members analyze youth and family member perspectives and stories to identify functional strengths.

1. National Wraparound Practice

Who is in the TOM national dataset?

What does the data say about ratings of fidelity as assessed in team meetings?

Method

- Data were collected by local evaluators or supervisors trained to criteria using the *TOM Training Toolkit*
- July 2009 to August 2012
- Uploaded into Wraparound Online Data Entry and Reporting System (WONDERS) and compiled in de-identified fashion by the research team

TOM Participants – Initial Sample

17 Projects

- *M* = 4.2 sites (*SD* = 7.5)
- Range 1 to 32 sites

72 Sites

- *M* = 19.5 meetings (*SD* = 20.9)
- Range 1 to 144 meetings

1,401 Team Mtgs

- Individual youth *n* = 1,304
- Initial meetings (18%)
- Follow-up (72%)
- Transition & “Other” (6%)

TOM Participants – Revised Sample

17 Projects

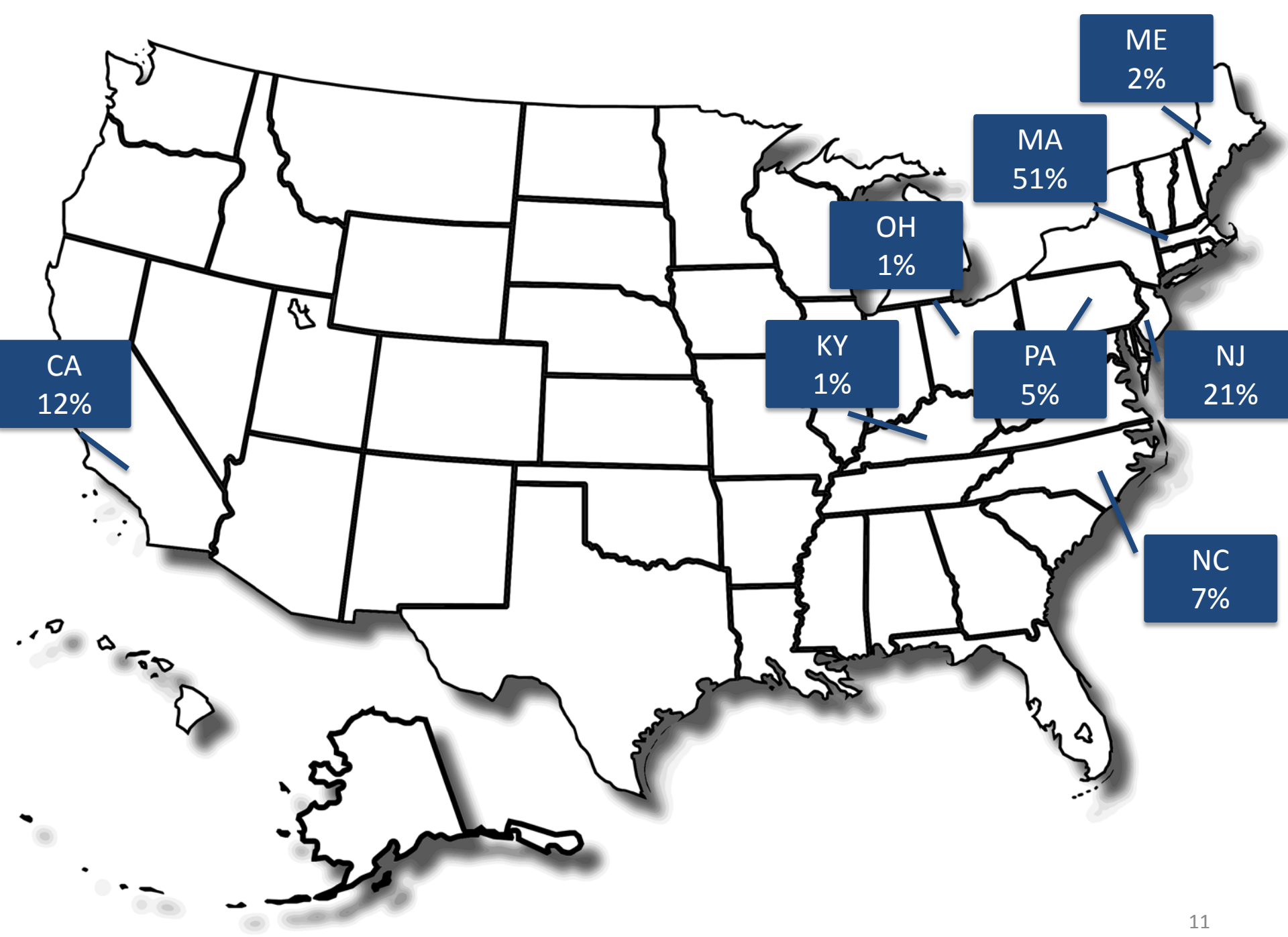
- *M* = 3.5 sites (*SD* = 7.0)
- Range 1 to 30 sites

59 Sites

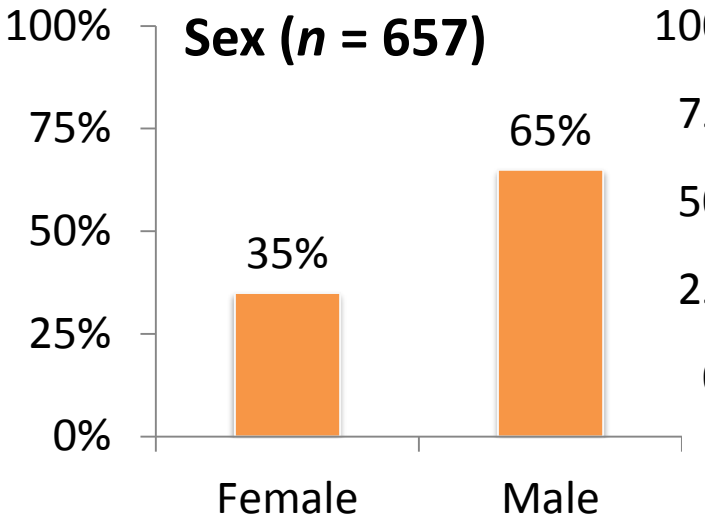
- *M* = 18.3 meetings (*SD* = 17.8)
- Range 5 to 129 meetings

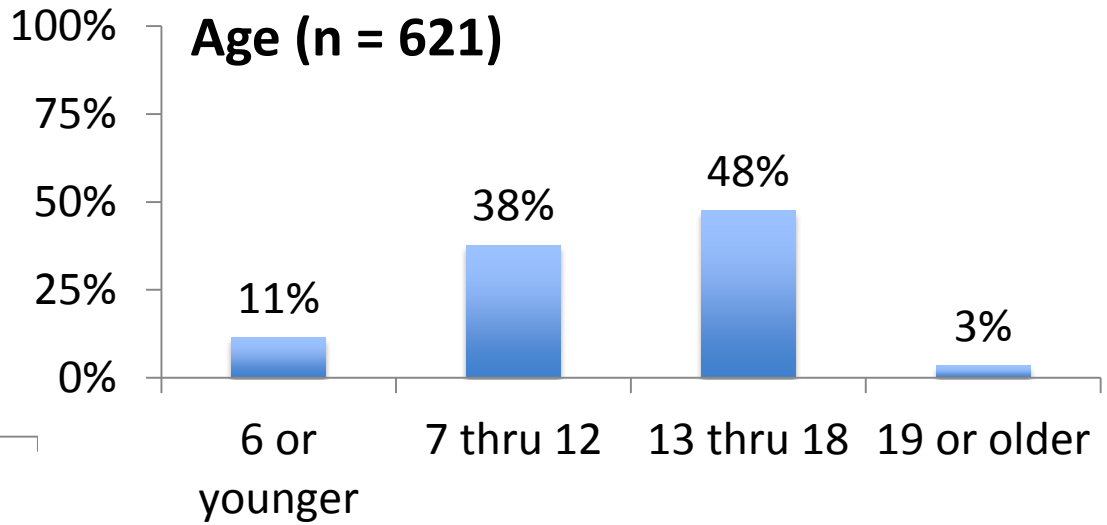
1,078 Team Meetings

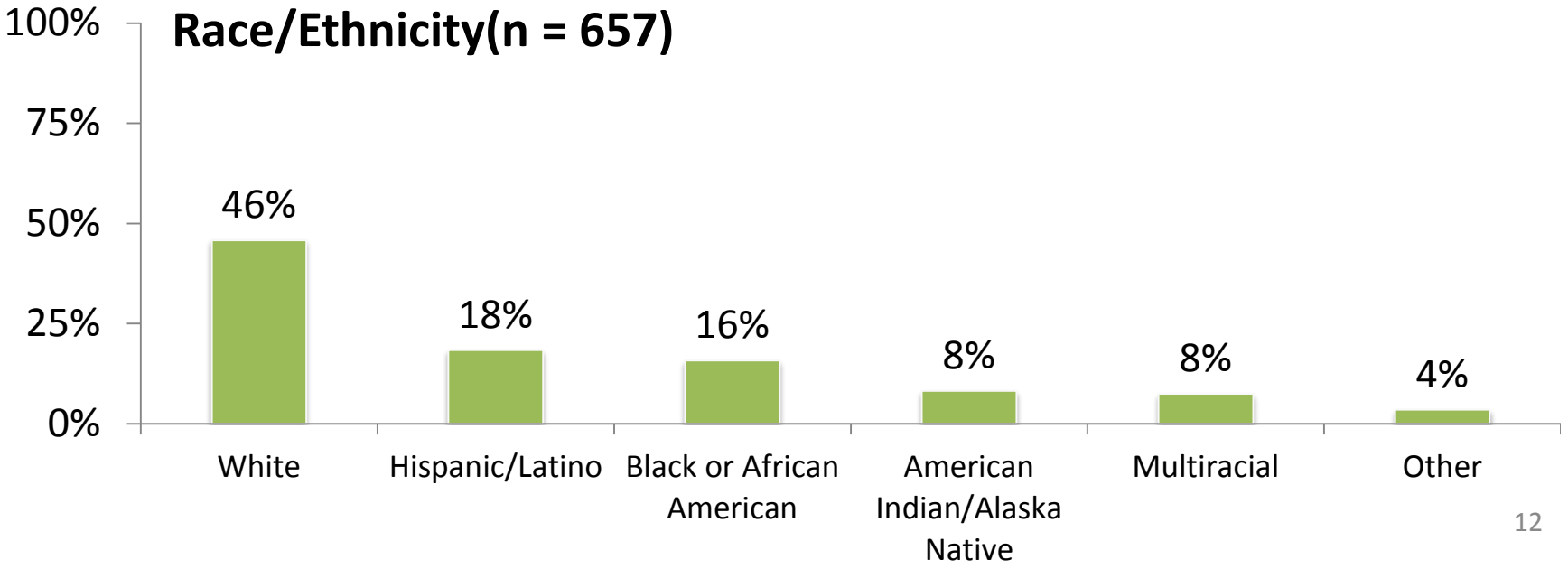
- Initial meetings (16%)
- Follow-up (76%)
- Transition (4%)


ME
2%
MA
51%
OH
1%
KY
1%
PA
5%
NJ
21%
CA
12%
NC
7%

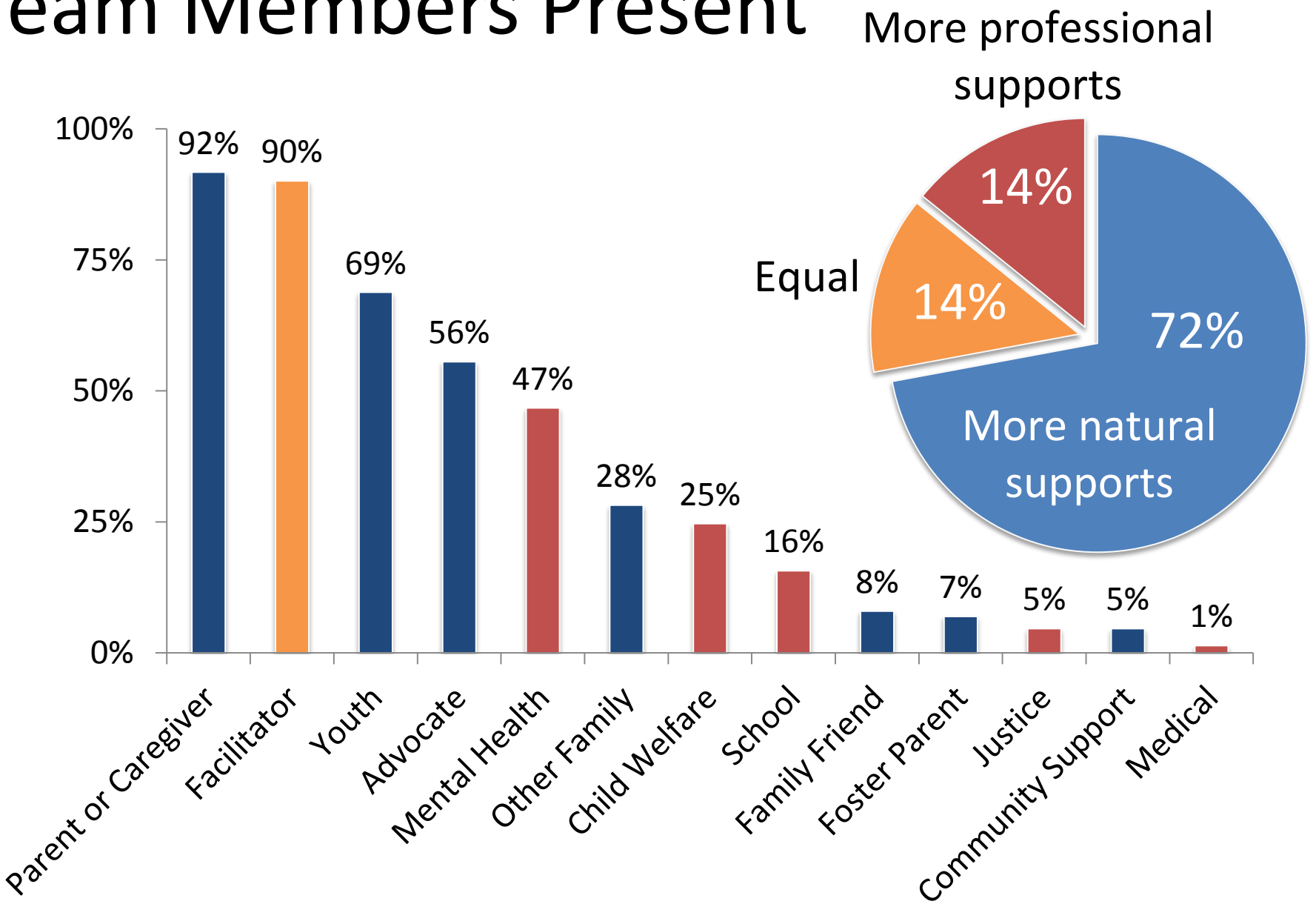
Youth Demographics (%)





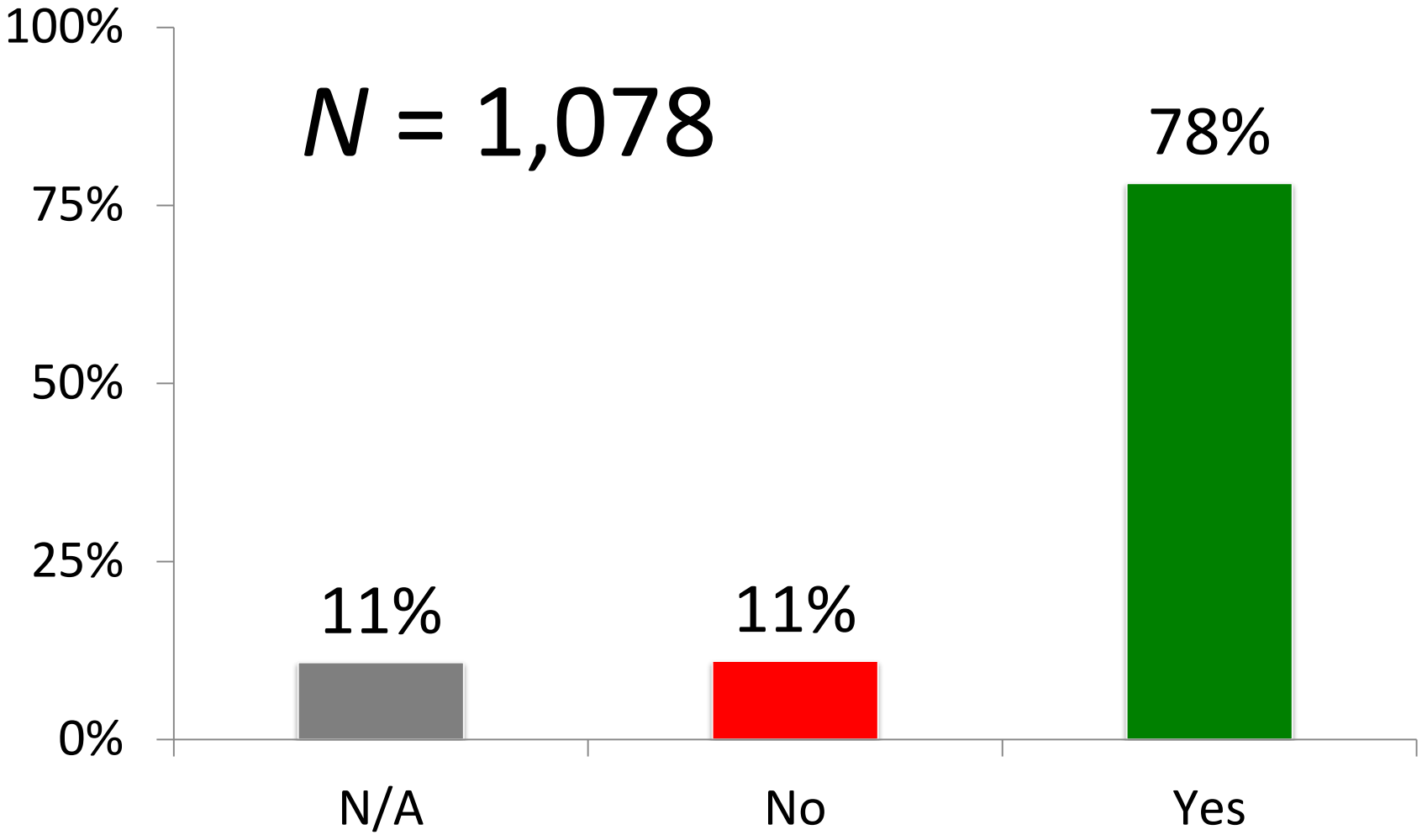


Team Members Present

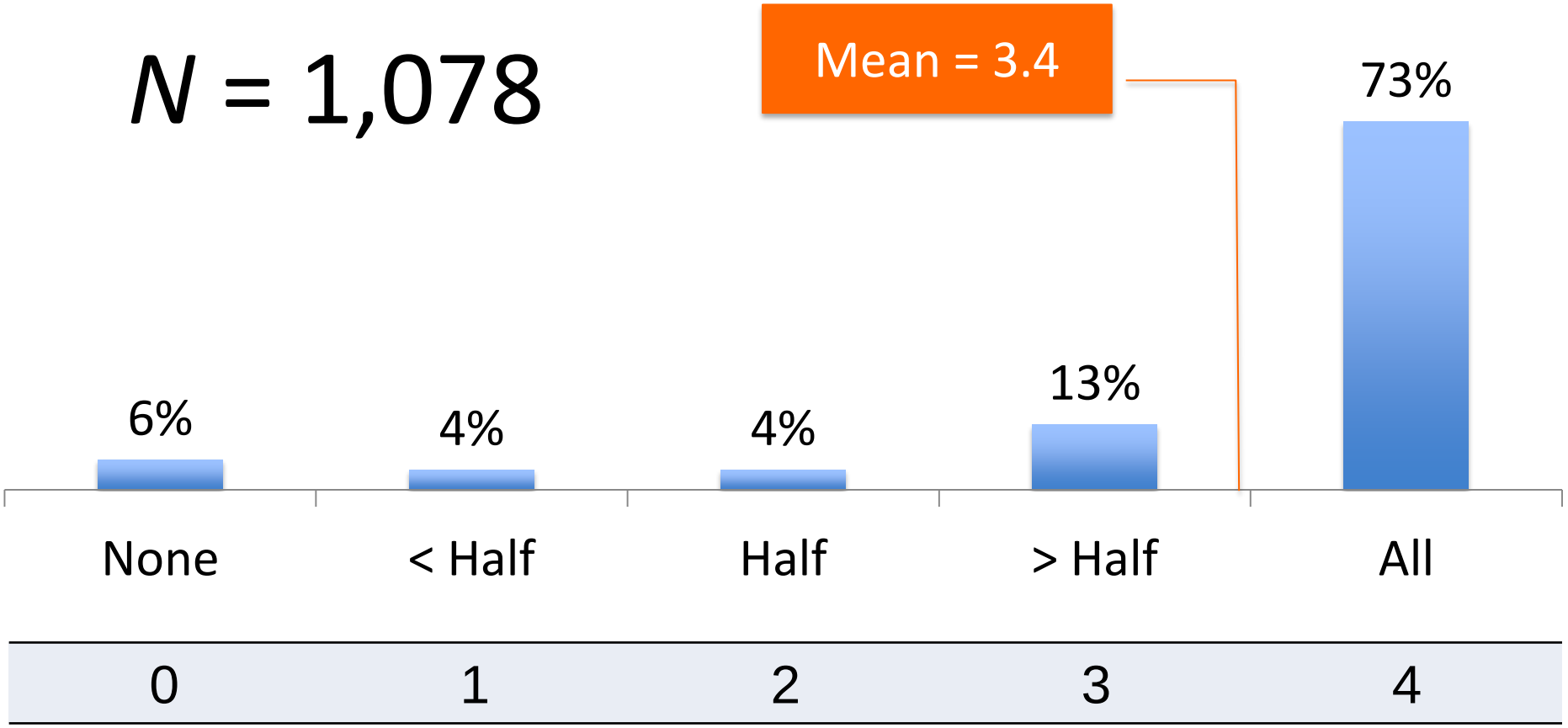


(n = 1,078, M = 6.1 SD = 2.2, 1 to 23)

TOM Indicator Responses



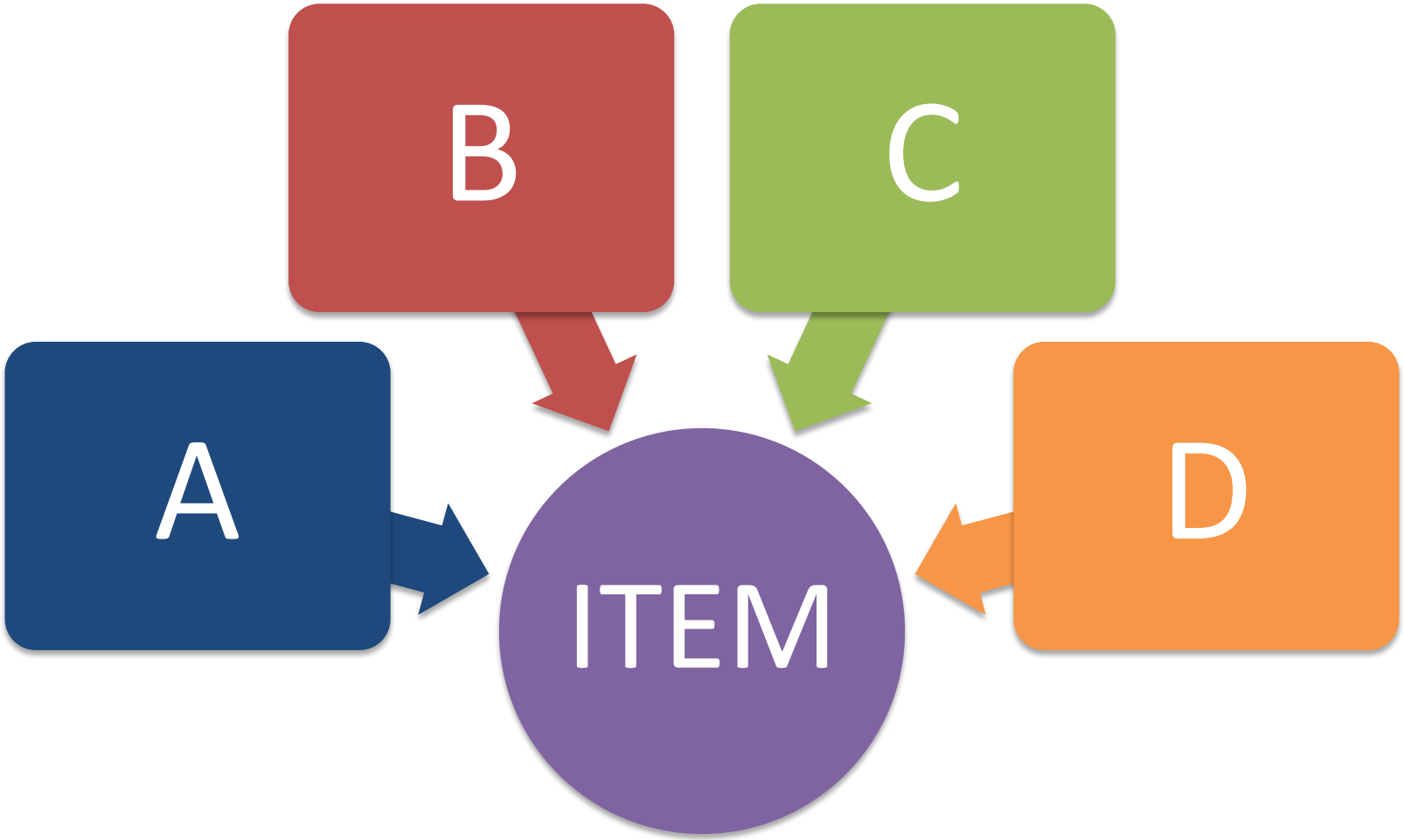
TOM Items: Indicators Present



2. Internal Consistency

How reliable is the TOM in terms of
(1) TOM Total scores and
(2) Item scores?

TOM Items = Sums of Indicators



TOM Items 1-10 (α = .80)

	Item	M	SD	α if deleted	Item Total *r*	α
1	Team Membership & Attendance	3.28	0.93	0.80	0.10	-
2	Effective Team Process	3.71	0.62	0.78	0.50	.42
3	Facilitator Preparation	3.51	0.89	0.79	0.35	.57
4	Effective Decision Making	3.66	0.69	0.78	0.43	.47
5	Creative Brainstorming & Options	3.30	1.33	0.78	0.45	.82
6	Individualized Process	3.70	0.64	0.78	0.47	.43
7	Natural and Community Supports	1.68	1.76	0.80	0.26	.90
8	Natural Support Plans	2.73	1.52	0.78	0.41	.50
9	Team Mission and Plans	3.68	0.67	0.79	0.34	.44
10	Shared Responsibility	3.71	0.79	0.78	0.43	.51

TOM Items 11-20 (α = .80)

	Item	M	SD	α if deleted	Item Total Cor	α
11	Facilitation Skills	3.28	0.93	0.78	0.53	.62
12	Cultural Linguistic Competence	3.71	0.62	0.79	0.36	.48
13	Outcomes Based Process	3.51	0.89	0.78	0.47	.78
14	Evaluating Progress and Success	3.66	0.69	0.78	0.52	.54
15	Youth and Family Voice	3.30	1.33	0.79	0.23	.64
16	Youth and Family Choice	3.70	0.64	0.79	0.32	.48
17	Focus on Strengths	1.68	1.76	0.78	0.45	.75
18	Positive Team Culture	2.73	1.52	0.78	0.48	.59
19	Community Focus	3.68	0.67	0.79	0.36	.71
20	Least Restrictive Environment	3.71	0.79	0.79	0.21	.65

3. Inter-rater Reliability

What is the inter-rater reliability of the TOM?

Does reliability vary by the type of observer?

Do TOM scores vary by type of observer?

Inter-rater reliability studies

- 2009 study of the initial version of the TOM conducted in California
  - 15 paired observations conducted by grad students
  - Pooled Kappa was .226 (fair agreement)
  - Results were used to revise the TOM, resulting in the current version of 71 indicators
- Two studies (2010 and 2012) have been conducted on the current TOM
  - One assessed reliability of two evaluators
  - One assessed reliability of a supervisor paired with an evaluator

Inter-rater Reliability Studies

- Pooled Kappa was used to assess agreement between raters in two studies.

- Pooled Kappa is the average of the averaged probabilities.

$$K_{pooled} = \frac{P_o - P_e}{1 - P_e}$$

- Differences in scoring patterns for two different types of TOM users were also examined.

Methods

2010 Study

Sample

- Paired raters attended 12 wraparound team meetings for 12 unique families in Nevada.

Method

- A research coordinator and wraparound program administrator were trained on the administration of the TOM and paired observations were conducted between October 2009 and February 2010.

2012 Study

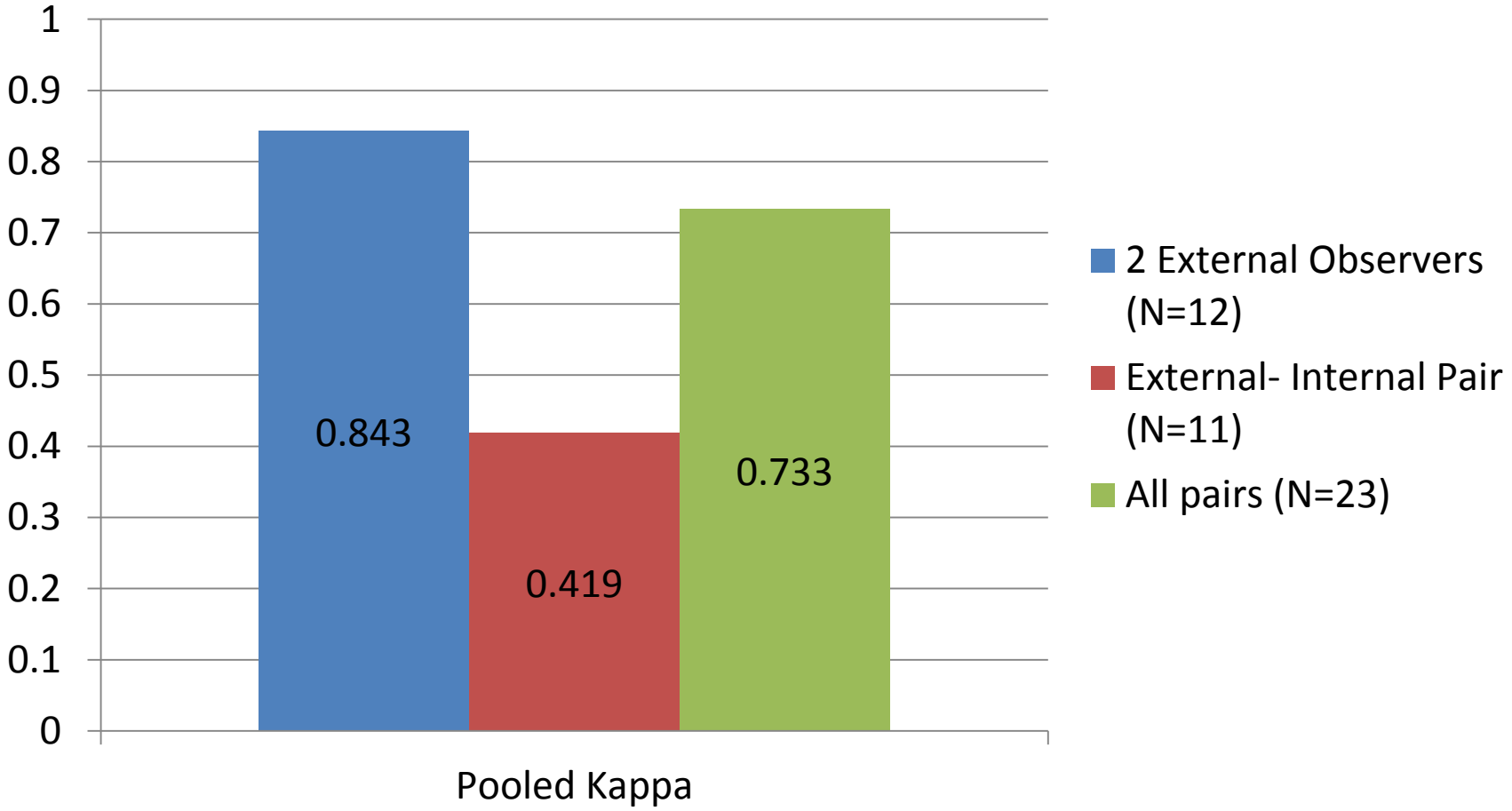
Sample

- Paired raters attended 11 wraparound team meetings for 11 unique families in Washington.

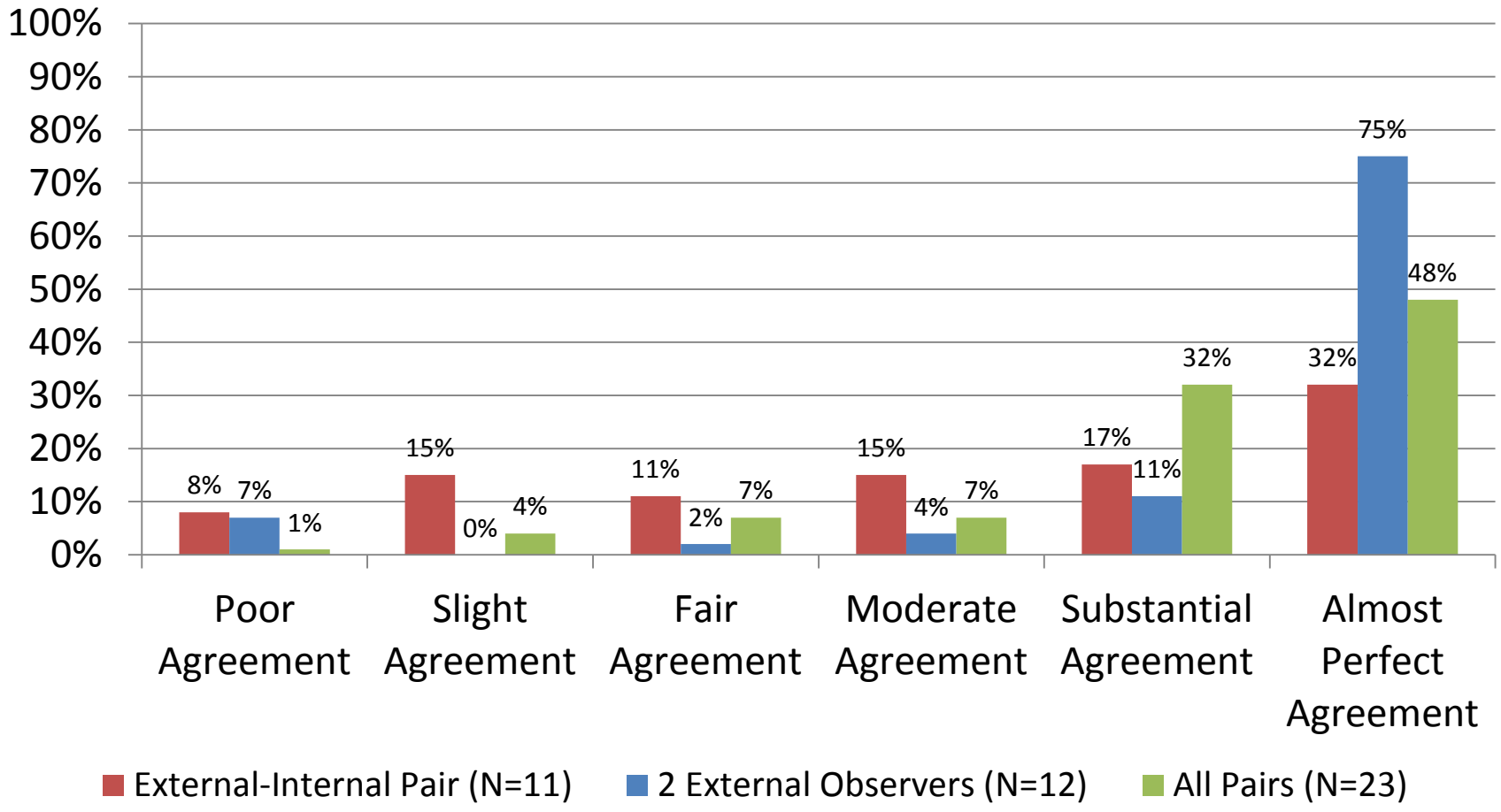
Method

- A researcher and wraparound coach were trained on the administration of the TOM and paired observations were conducted between April and August 2012.

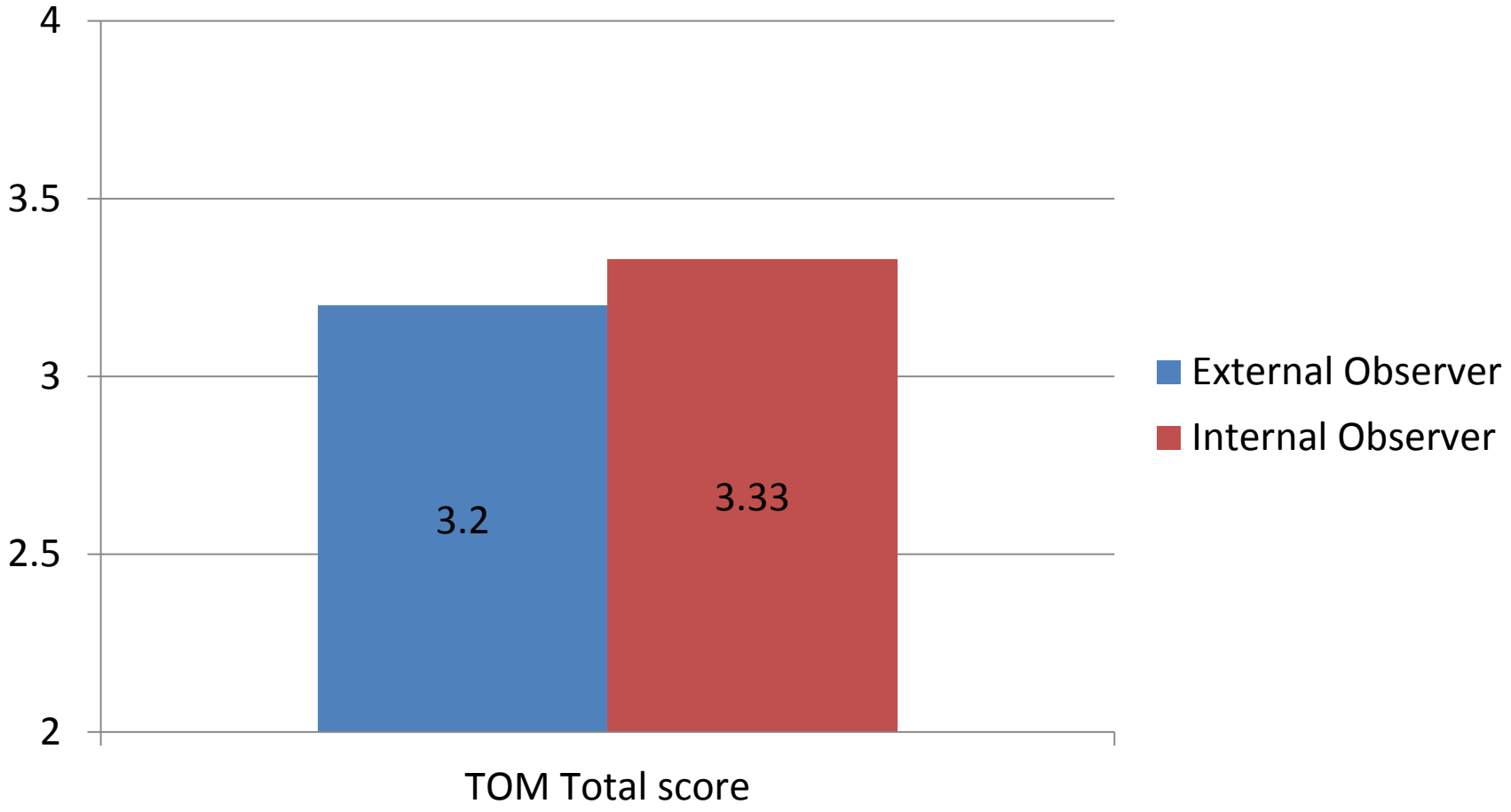
Results: Agreement was higher when 2 external observers observed teams



Percent of TOM Indicators Showing Different Levels of Agreement by Type of Rater Pair



TOM Mean score was higher for internal observers than external observers in Washington



Differences in Scoring Patterns by Rater Type in Washington

Item	External Observers Mean	Internal Observers Mean
Item 1: Team Membership and Attendance	3.18	**3.64**
Item 2: Effective Team Process	3.18	**3.55**
Item 3: Facilitator Preparation	3.27	**3.45**
Item 4: Effective Decision Making	**3.09**	2.82
Item 5: Creative Brainstorming and Options	2.91	**3.27**
Item 6: Individualized process	4.00	4.00
Item 7: Natural and Community Supports	1.45	**1.82**
Item 8: Natural Support Plans	2.82	**3.36**
Item 9: Team Mission and Plans	3.45	3.45
Item 10: Shared Responsibility	3.55	**3.73**
Item 11: Facilitation Skills	**3.36**	3.27
Item 12: Cultural and Linguistic Competence	3.82	3.82
Item 13: Outcomes Based Process	2.44	**2.78**
Item 14: Evaluating Progress and Success	**3.18**	2.64
Item 15: Youth and Family Voice	3.18	**3.73**
Item 16: Youth and Family Choice	**3.55**	3.40
Item 17: Focus on Strengths	2.91	2.91
Item 18: Positive Team Culture	**3.55**	3.09
Item 19: Community Focus	3.36	**3.91**
Item 20: Least Restrictive Environment	4.00	4.00
TOTAL TOM MEAN SCORE	3.20	3.33

Findings and Conclusions

- Substantial agreement between raters overall
  - Provides evidence of TOM inter-rater reliability.
  - This is important because IRR is probably the most important type of reliability for observation tools
- Almost perfect agreement for pair of external observers; only moderate agreement for external observers paired with internal observers.
  - This difference could possibly be attributed to rater type.
  - Supervisors also rated teams as showing higher fidelity
  - TOM use by supervisors may be cost-effective and aid feedback to staff, but one may question validity of results
- The TOM will be revised based on results of these studies.
  - Inter-rater reliability by indicator will be used to remove or revise some indicators

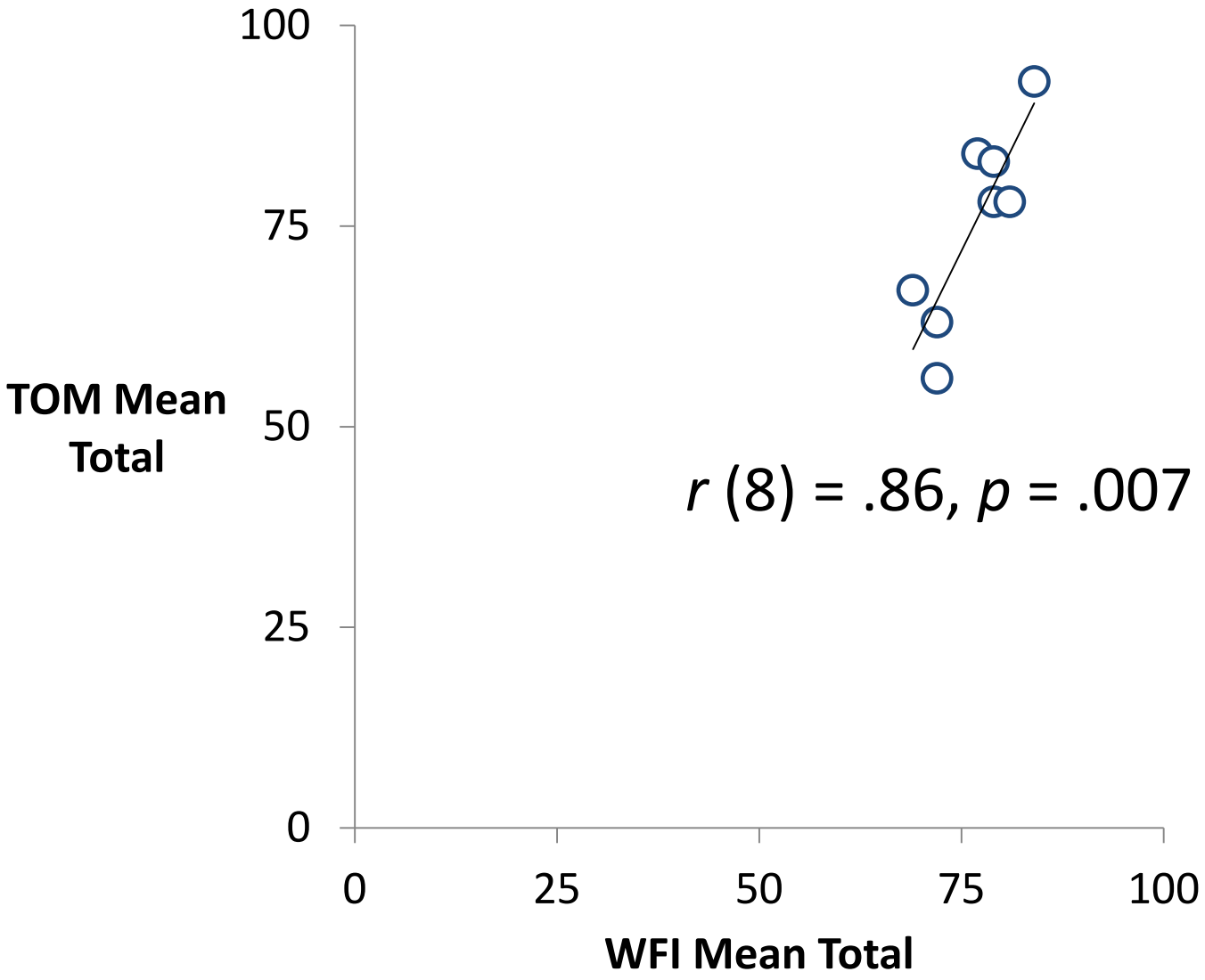
4. Concurrent Validity

Does the TOM show validity in the form of correlation with another wraparound fidelity tool (the WFI-4)?

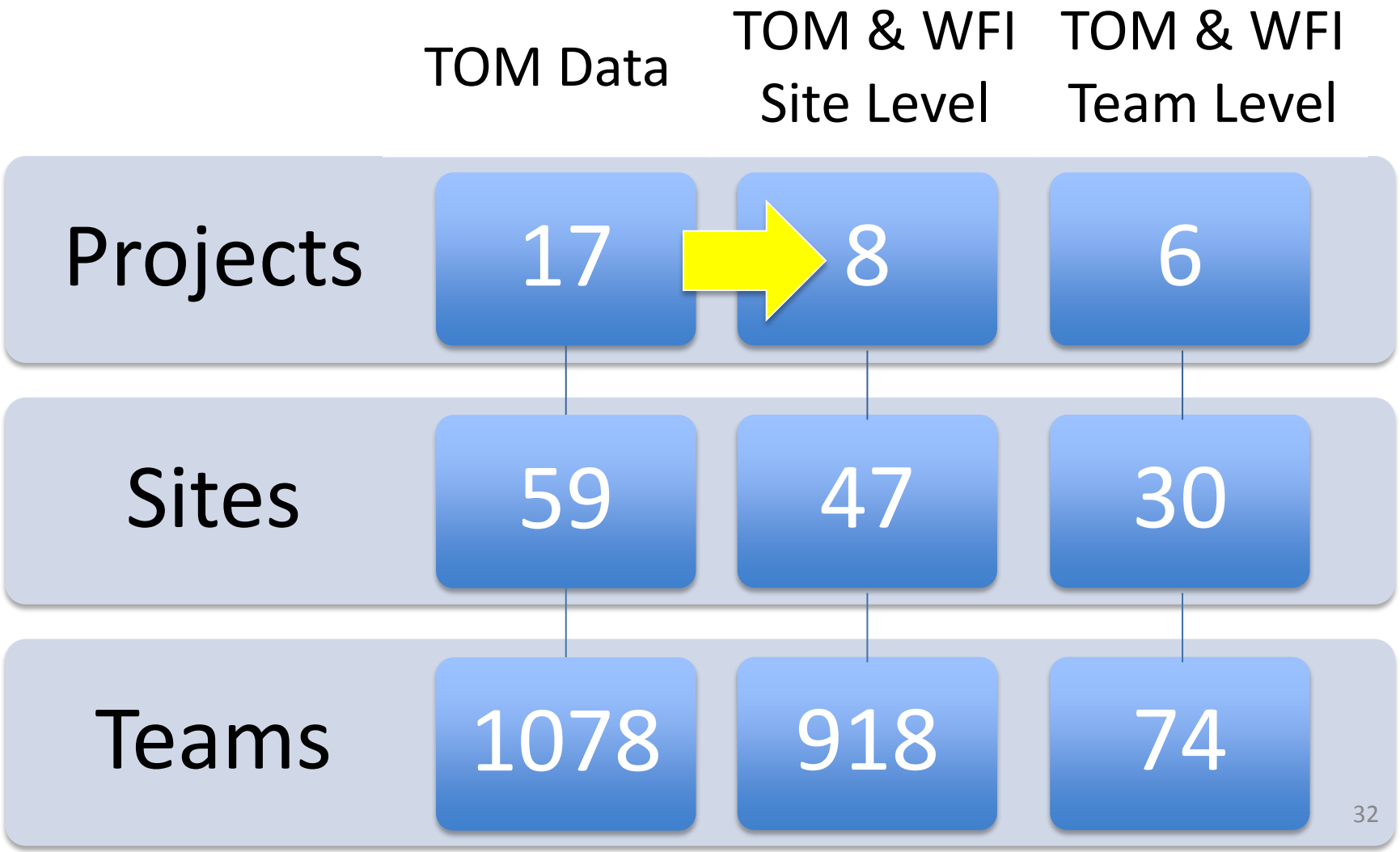
Concurrent Validity Study

- Comparison between TOM and WFI fidelity
- Validation of TOM and provide guidance to users of both measures
- Previous study showed strong relationship at project level for using earlier version of TOM
- Current study uses revised version of TOM with comparisons at multiple levels

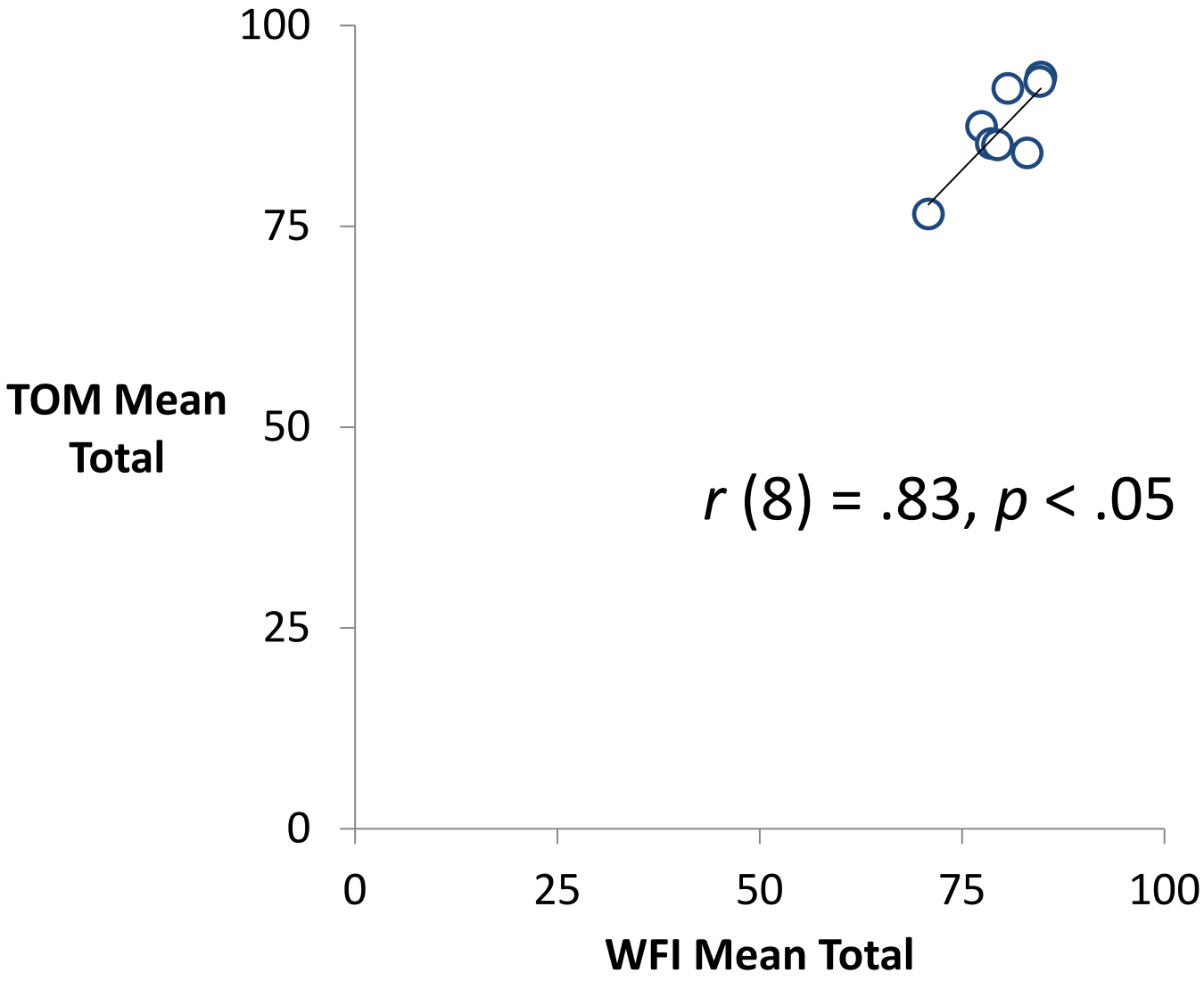
Previous Study: 8 Sites in Same Project



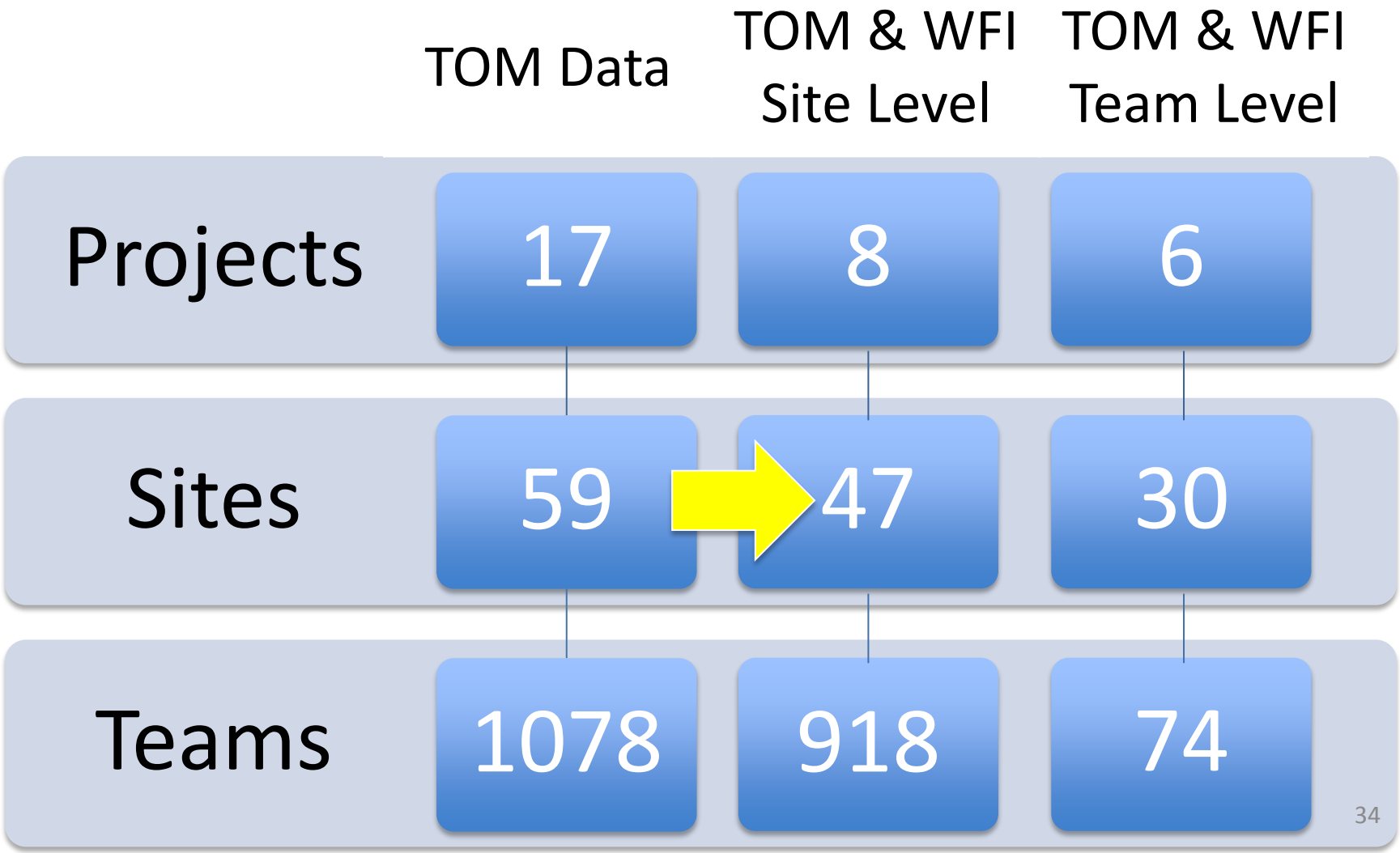
Nested Sample



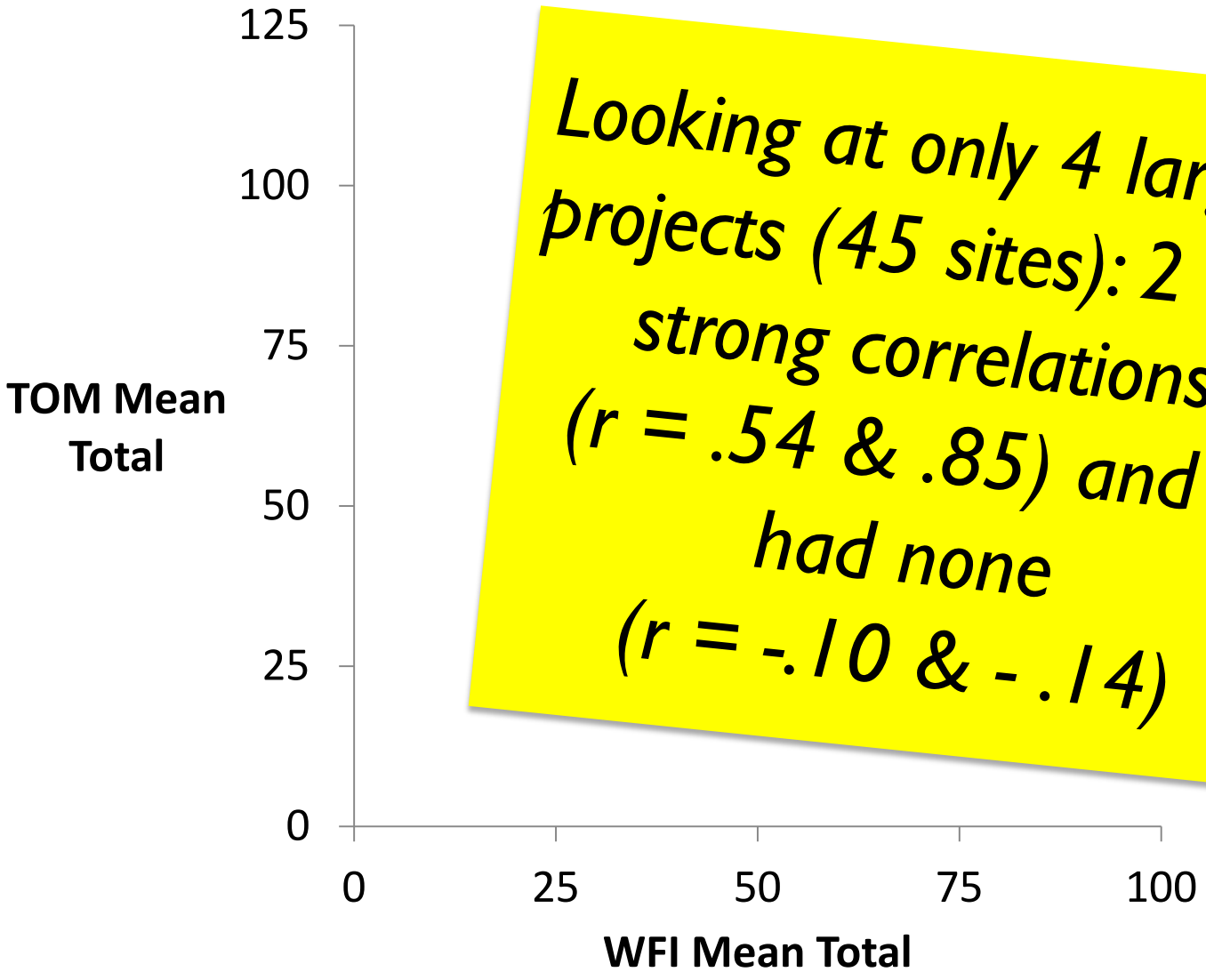
Project Level Correlation



Nested Sample

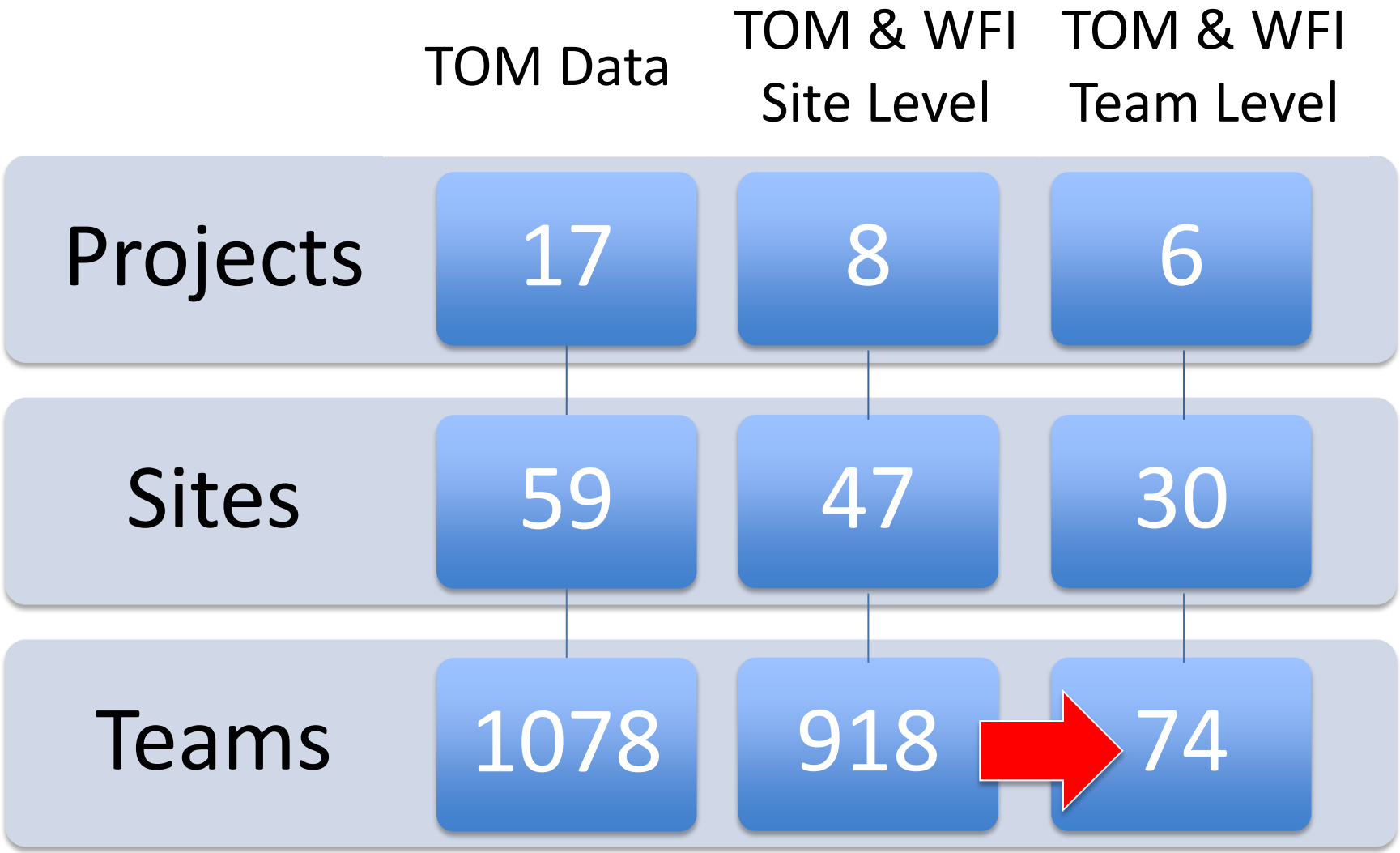


Site Level Correlation

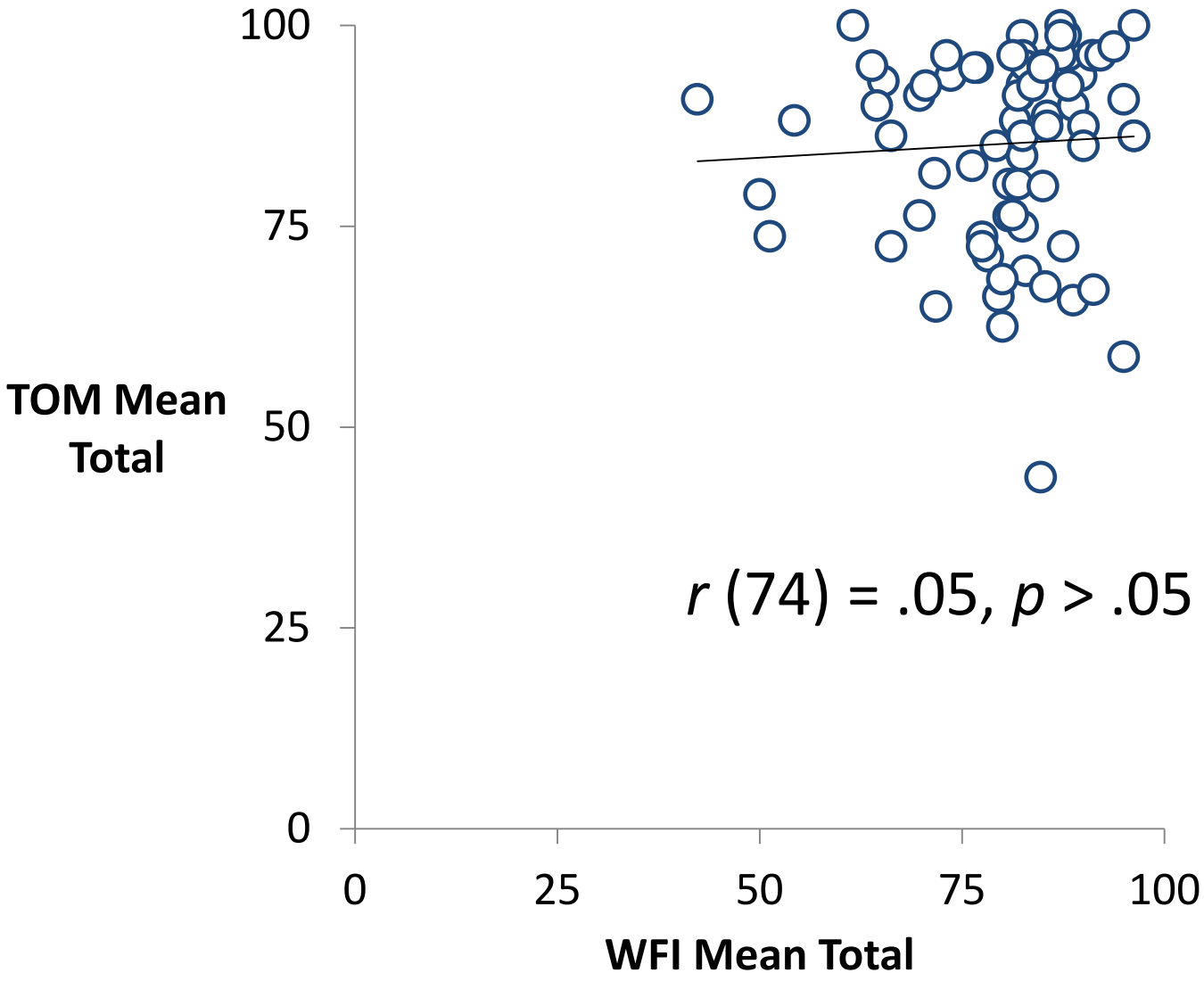


Looking at only 4 largest projects (45 sites): 2 had strong correlations (r = .54 & .85) and 2 had none (r = -.10 & - .14)

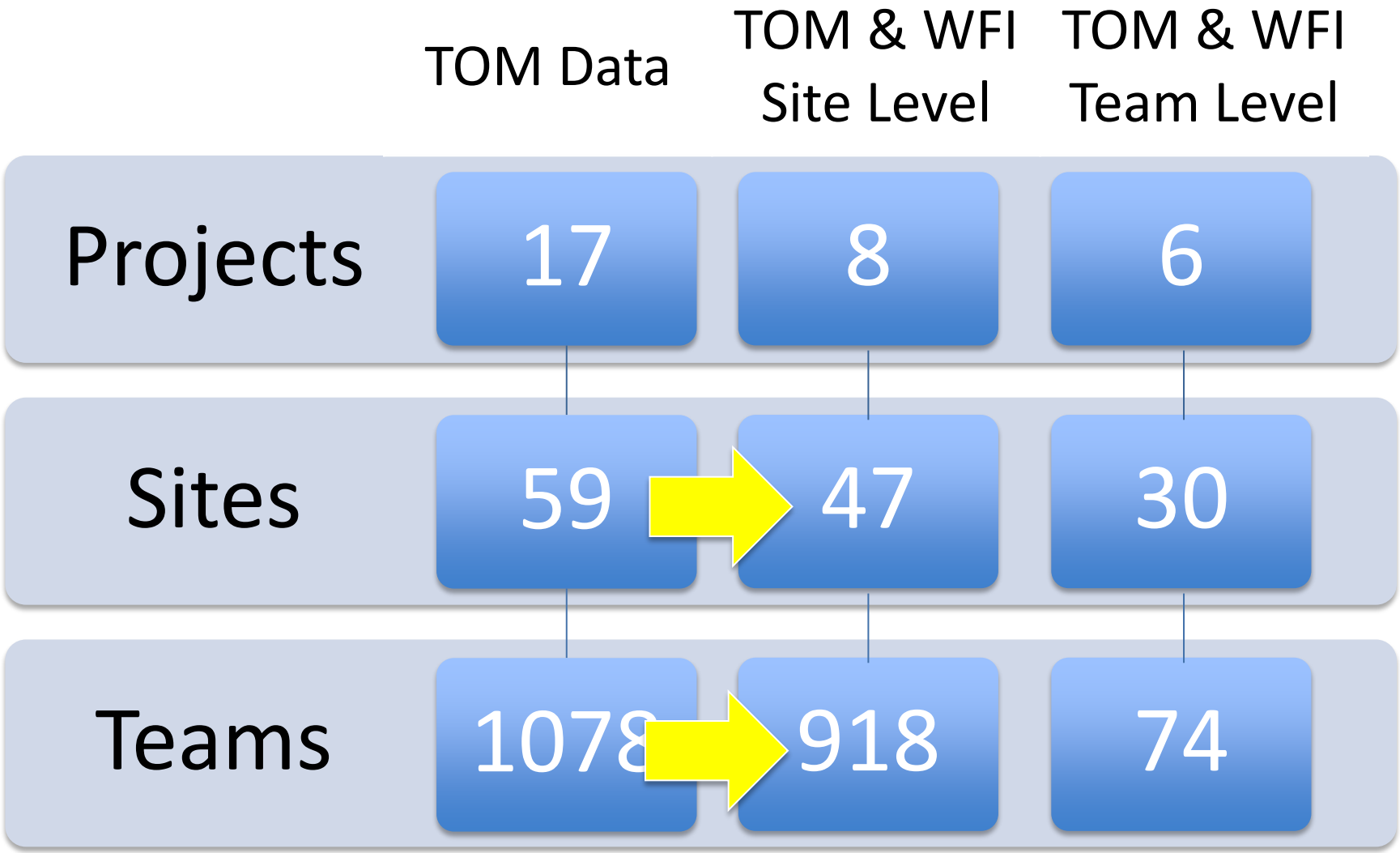
Current Study - Nested Sample



Team Level Correlation



Current Study - Nested Sample



Multi-Level Modeling

Parameter	Model 1	Model 2	Model 3
Intercept		-1.36 (.95)	-0.71 (1.06)
Level 1 (Team)			
Follow up meeting			.83 (.96)
Level 2 (Site)			
WFI Rating			.01 (.29)
External Observer			10.93 (7.1)
Random Effects			
Residual			111.78 (5.73)*
Intercept			28.35 (8.48)*
Model Statistics			
AIC	7342.19	6322.82	6132.56
BIC	7351.91	6332.26	6141.93

No relationship found between TOM (team level) & WFI (site level) using HLM

Findings

- Very high endorsement of wraparound indicators
  - TOM and WFI scores increasing nationally every year
  - Reduces utility as a research an QA tool
- Stronger relationships at higher levels
  - TOM and WFI function as a valid agency or program level assessment
  - TOM and WFI may tap into very different things at an individual team/family level
- Site level WFI-TOM correlation low overall, but:
  - Very strong for projects using external evaluators
  - Very poor for projects using supervisors

Overall implications