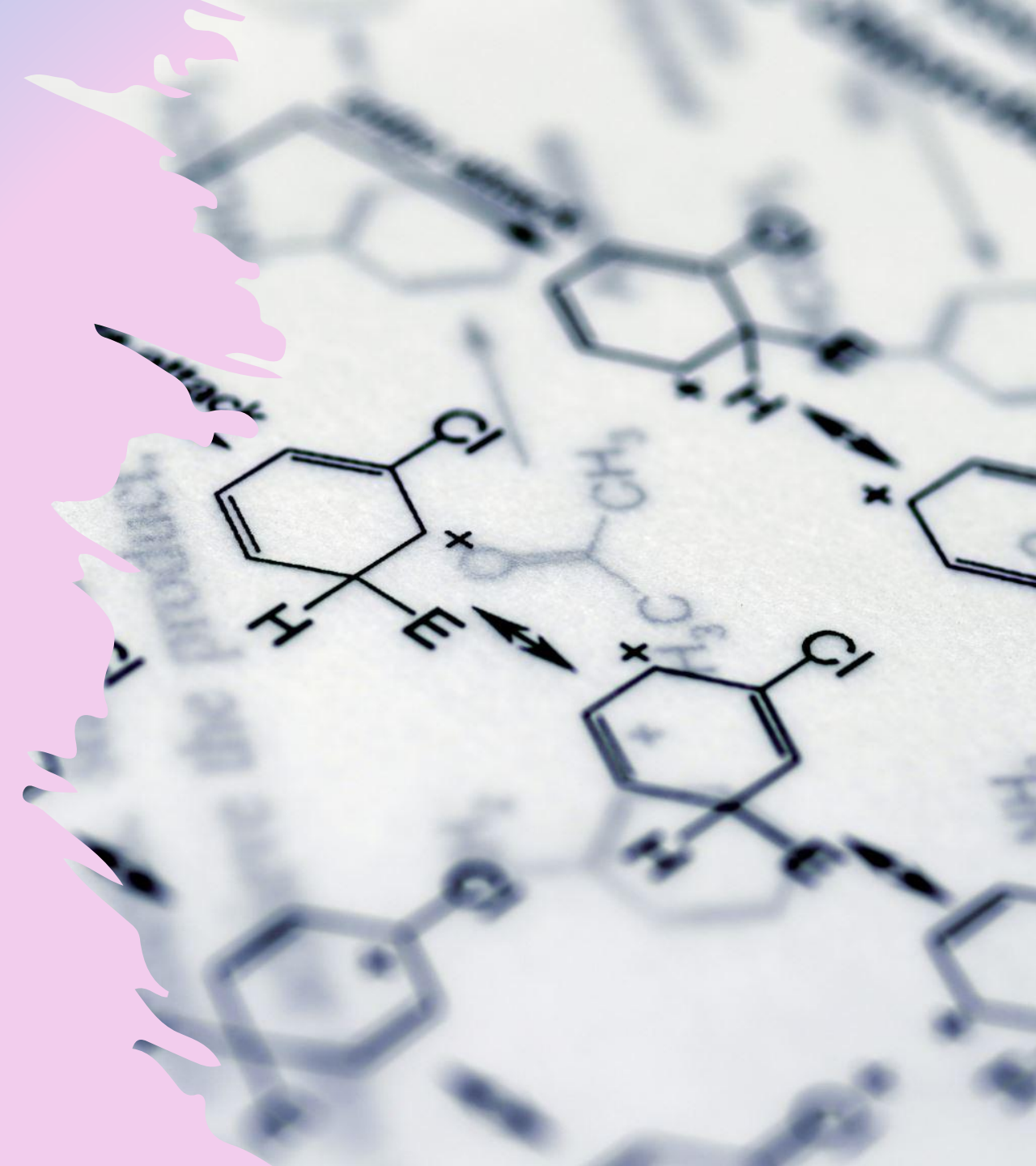# Interobserver Agreement: A Preliminary Investigation into How Much Is Enough?

- Hausman, N. L., Javed, N., Bednar, M. K., Guell, M., Schaller, E., Nevill, R. E., & Kahng, S.

Abstract

IOA ensures data reliability and validity by confirming that observations are consistent and unbiased.

Current standards for acceptable IOA levels are subjective and vary across studies and settings.

This study examines variables influencing IOA sufficiency and offers recommendations for empirically grounded benchmarks.

Introduction

IOA is a cornerstone of reliable data collection in research and practice, ensuring observations are free from observer bias.

Existing benchmarks, such as 80% or 90% agreement, are widely accepted but lack empirical validation.

This study seeks to fill this gap by exploring factors that affect IOA sufficiency and proposing data-driven benchmarks.
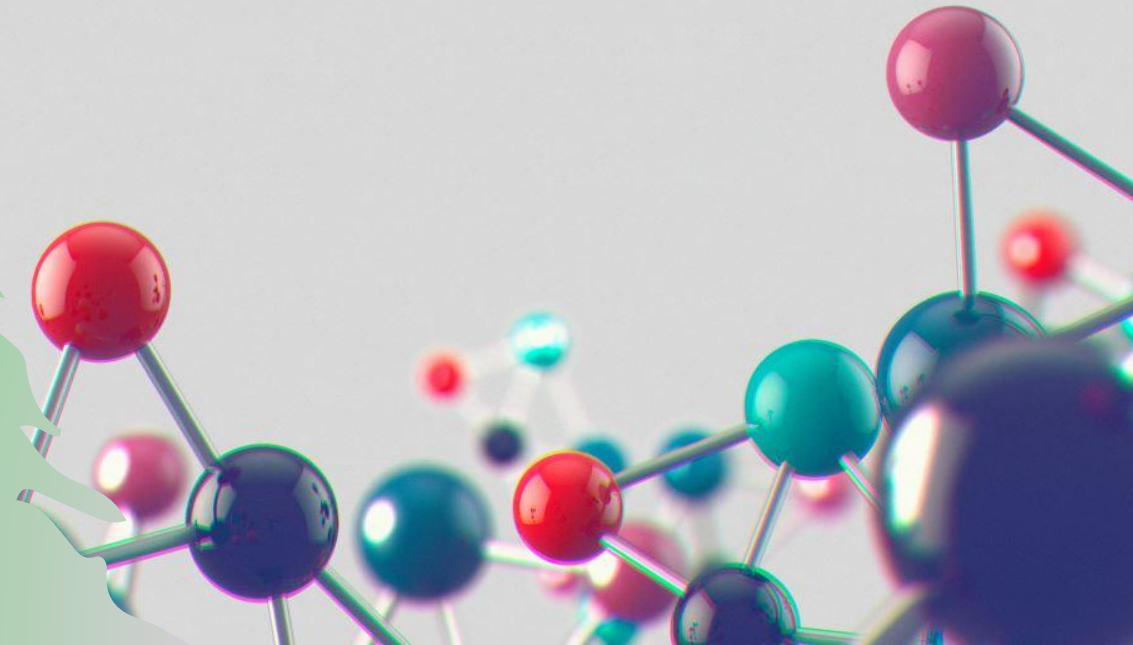
Research Objectives

Explore
- Explore the factors influencing IOA sufficiency, such as behavior complexity, observer experience, and observational conditions.

Assess
- Assess the impact of different data characteristics, including frequency, duration, and latency, on IOA.

Develop
- Develop evidence-based recommendations for IOA benchmarks that are adaptable to specific contexts.

Methodology - Participants

Observers included both novice and experienced professionals, such as RBTs and BCBAs.

Behaviors targeted for observation encompassed social interactions, academic skills, and motor actions, ensuring a broad range of data.
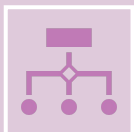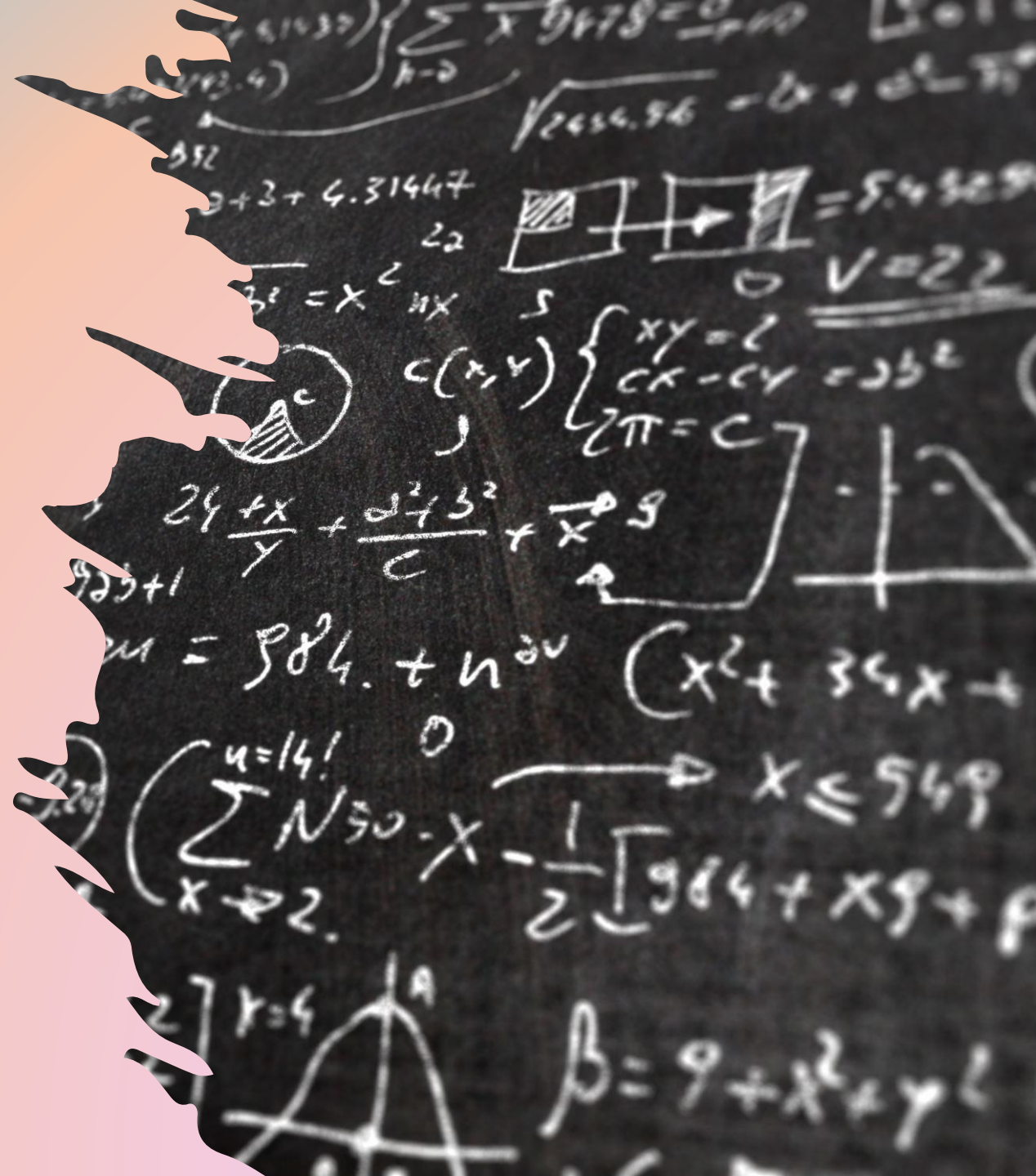
Methodology - Data Collection

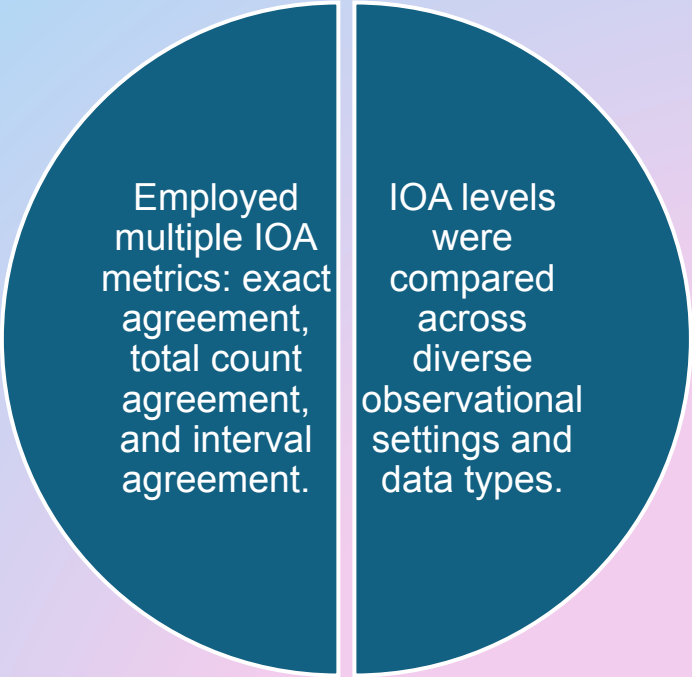Sessions were recorded to allow for repeated and independent observations.

Observers scored data on three dimensions: frequency, duration, and latency of behaviors.

Multiple conditions were tested, including structured and unstructured environments.

Methodology - IOA Calculation

Employed multiple IOA metrics: exact agreement, total count agreement, and interval agreement.

IOA levels were compared across diverse observational settings and data types.

Methods of Calculating IOA - Exact Agreement

Definition: Exact agreement occurs when observers record identical scores for each data point.

Formula: (Number of exact agreements / Total number of intervals) x 100.

Strengths: High level of precision.

Limitations: Difficult to achieve with complex behaviors.

Methods of Calculating IOA - Total Count Agreement

Definition: Compares the total counts recorded by each observer.

Formula: (Smaller total / Larger total) x 100.

Strengths: Simpler calculation.

Limitations: Does not account for distribution of behaviors across intervals.

Methods of Calculating IOA - Interval-by-Interval Agreement

Definition: Compares observer records within each interval.

Formula: (Number of agreements / Total intervals) x 100.

Strengths: Accounts for behavior distribution.

Limitations: Time-intensive for larger datasets.

Results – Overview

IOA levels varied significantly depending on the type of data and the conditions under which observations were made.

Structured settings yielded higher IOA compared to unstructured settings, demonstrating the impact of environmental control.

Results - Data Characteristics

Simpler behaviors, such as discrete motor actions, yielded higher IOA.

Complex or multi-faceted behaviors showed greater variability in agreement.

Results – Observer Experience

Experienced observers consistently achieved higher IOA compared to novices.

Targeted training programs improved IOA among less experienced observers.

Discussion

Fixed IOA benchmarks may not be appropriate across all contexts.

Contextual factors, such as behavior complexity and observer training, must be considered.
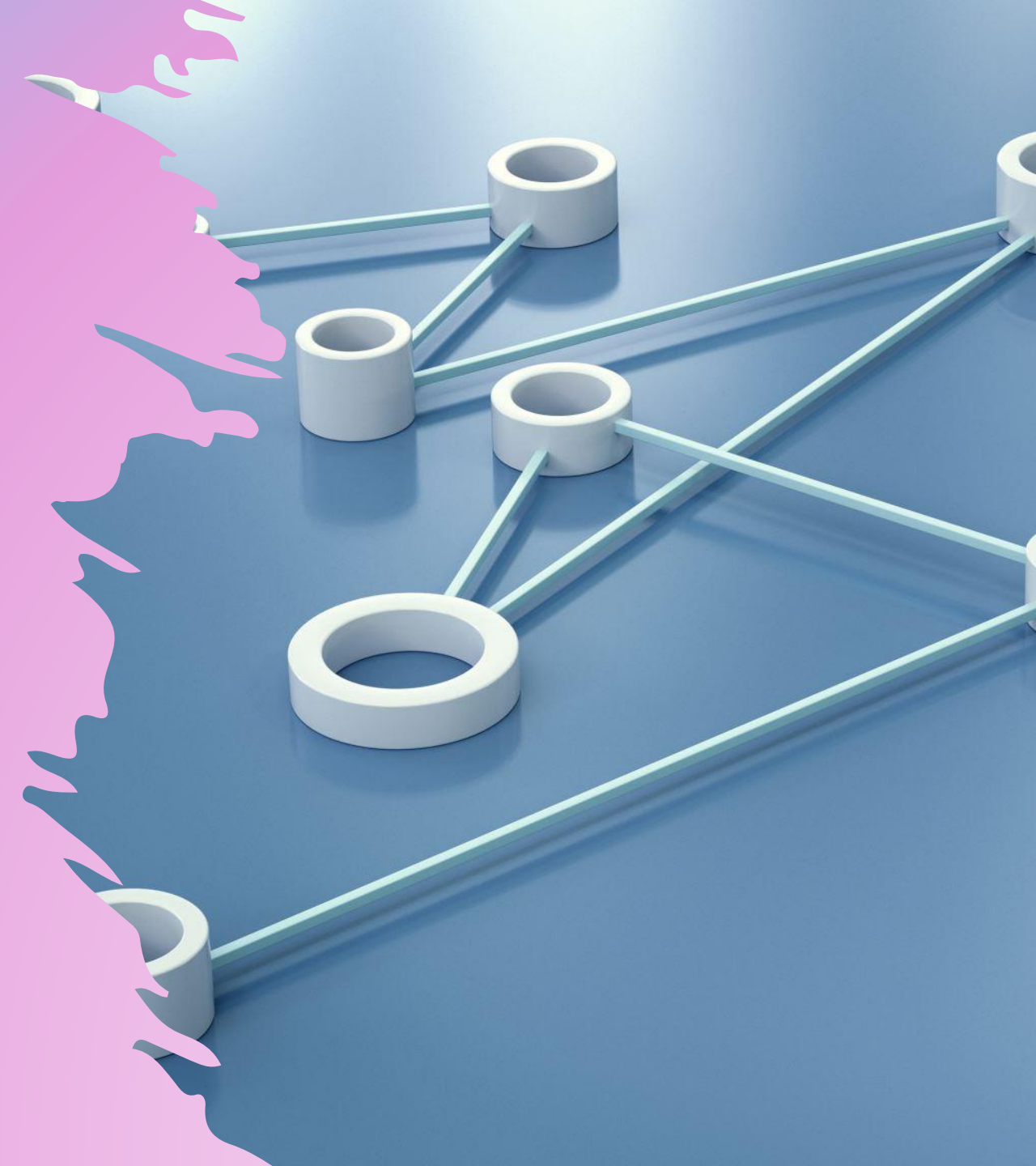
Recommendations

Develop adaptive IOA benchmarks tailored to specific data and settings.

Emphasize comprehensive observer training to enhance IOA consistency.

Limitations

Limited generalizability due to a controlled study environment.

Small sample size may restrict broader application of findings.

Future Directions

Investigate IOA benchmarks in real-world settings.

Examine the long-term effects of observer training on IOA.

Conclusion

IOA sufficiency is highly context-dependent.

Evidence-based benchmarks are crucial for reliable data.

References

- Hausman, N. L., Javed, N., Bednar, M. K., Guell, M., Schaller, E., Nevill, R. E., & Kahng, S. (2022). Interobserver agreement: A preliminary investigation into how much is enough? *Journal of Applied Behavior Analysis, 55*(2), 357-368. https://doi.org/10.1002/jaba.811