

## INTRODUCTION

- Treatment integrity (TI) is the degree to which a prescribed intervention is implemented as intended (Gresham et al., 2000)
- Behavior analysts collect TI data to:
  - (a) assess the quality of implementation
  - (b) guide programming decisions
  - (c) train behavior-change agents
- Purpose: To compare various methods of TI data collection for their reliability, specificity, and completion time

## METHOD

**Participants:** 3-year-old female with autism spectrum disorder and four Registered Behavior Technicians

**Setting:** University-based autism treatment center

### Materials:

- 10 video samples (~150 trials) of 1:1 discrete trial instruction to teach noun and verb tacts
- Data sheets for scoring types
- Timer
- Operational definitions, session protocol, and challenging behavior definitions

**General procedure:** Remote scoring of video samples by trained research assistants using video software and electronic TI data sheets

- Only program protocols and operational definitions accessible
- No note-taking; only recording what the data sheet permits
- Pause and rewind permitted at any time

## METHOD

### Scoring and Comparison

Independent data collectors scored the same video samples using different TI measures (see 1-3)

- Data collectors were assigned 2-3 different scoring types from different categories (e.g., Likert by Trial & All-or-None by Component)
- Data collectors were permitted to score the same video twice using different methods if 5 videos were scored in between re-scoring

#### Interobserver Agreement

- Exact Interobserver Agreement calculated for all scoring types using: (# of exact agreements on trials or components scored / total # of trials or components scored) X 100
- Interobserver Agreement calculated for time to score using: (shorter duration / longer duration) X 100

Data collection by component

Teaching Session		
Components	Trial 1	Trial 2
Attending	+ / -	+ / -
Present Materials	+ / -	+ / -
Instruction	+ / -	+ / -
Social Reinforcer	+ / -	+ / -
Tangible Reinforcer	+ / -	+ / -
Prompt	+ / -	+ / -
Remove Materials	+ / -	+ / -
Record Data	+ / -	+ / -
Inter-trial Interval	+ / -	+ / -
Challenging Behavior	+ / -	+ / -

Data collection by trial

### 1 Occurrence / Non-Occurrence

Data collectors score the occurrence or non-occurrence of components across all trials in the session.

Score	Occurrence / Non-Occurrence by Component
+	Component implemented correctly
-	Error of omission and/or commission on component

Data collectors calculate integrity by component, by trial, and for the overall session

- Percent correct implementation for trials and components is calculated using: (# of trials or components implemented correctly / total number of trials or components scored) X 100
- An overall integrity score is calculated using: ( trials implemented correctly / total # of trials) X 100

*Adapted from Carroll et al. (2013)*

### 2 All-or-None

Data collectors score whether all components or trials are implemented correctly in the session.

Score	By Trial	By Component
+	All components were implemented correctly on this trial	This component was implemented correctly on every trial
-	At least one error occurred on this trial	This component was implemented incorrectly at least once in the session
(N/A)	---	This component was not applicable to this session

Data collectors calculate overall session percent integrity

- (# of trials or components implemented correctly 100% of the time / total number of trials or components scored) X 100

### 3 Likert Scales

Data collectors rate the implementation of components or trials in the session.

Score	5-Point (By Trial or Component)	3-Point (By Trial or Component)
1	Provider never implements appropriately (0%)	Provider does not implement during the session or never implements appropriately.
2	Provider occasionally implements steps competently, but misses many steps (1-49%)	Provider implements steps competently occasionally but misses many opportunities. Provider implements competently half of the time but misses many opportunities.
3	Provider implements half of the steps competently, but misses many steps (50-79%)	Provider implements steps competently most of the time but misses some opportunities. Provider implements competently throughout session.
4	Provider implements steps competently most of the time, but misses some steps (80-99%)	---
5	Provider implements steps competently throughout whole session (100%)	---

Data collectors average Likert scales to produce one overall Likert score using

- (sum of scores across trials or components / # of trials or components scored)

*Adapted from Suhrheinrich et al. (2019)*

## RESULTS

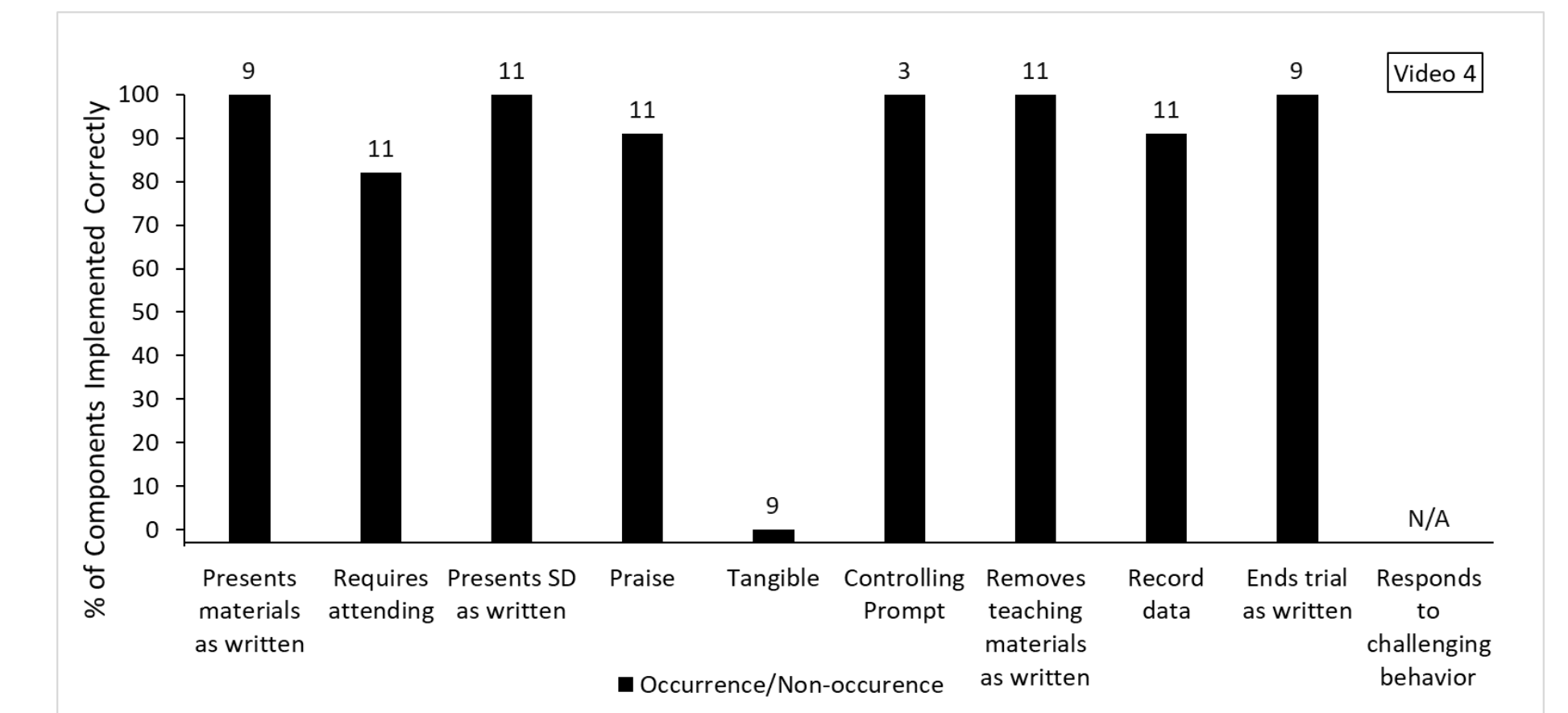


Figure 1. Percentage of components implemented correctly in one video sample (Video 4).

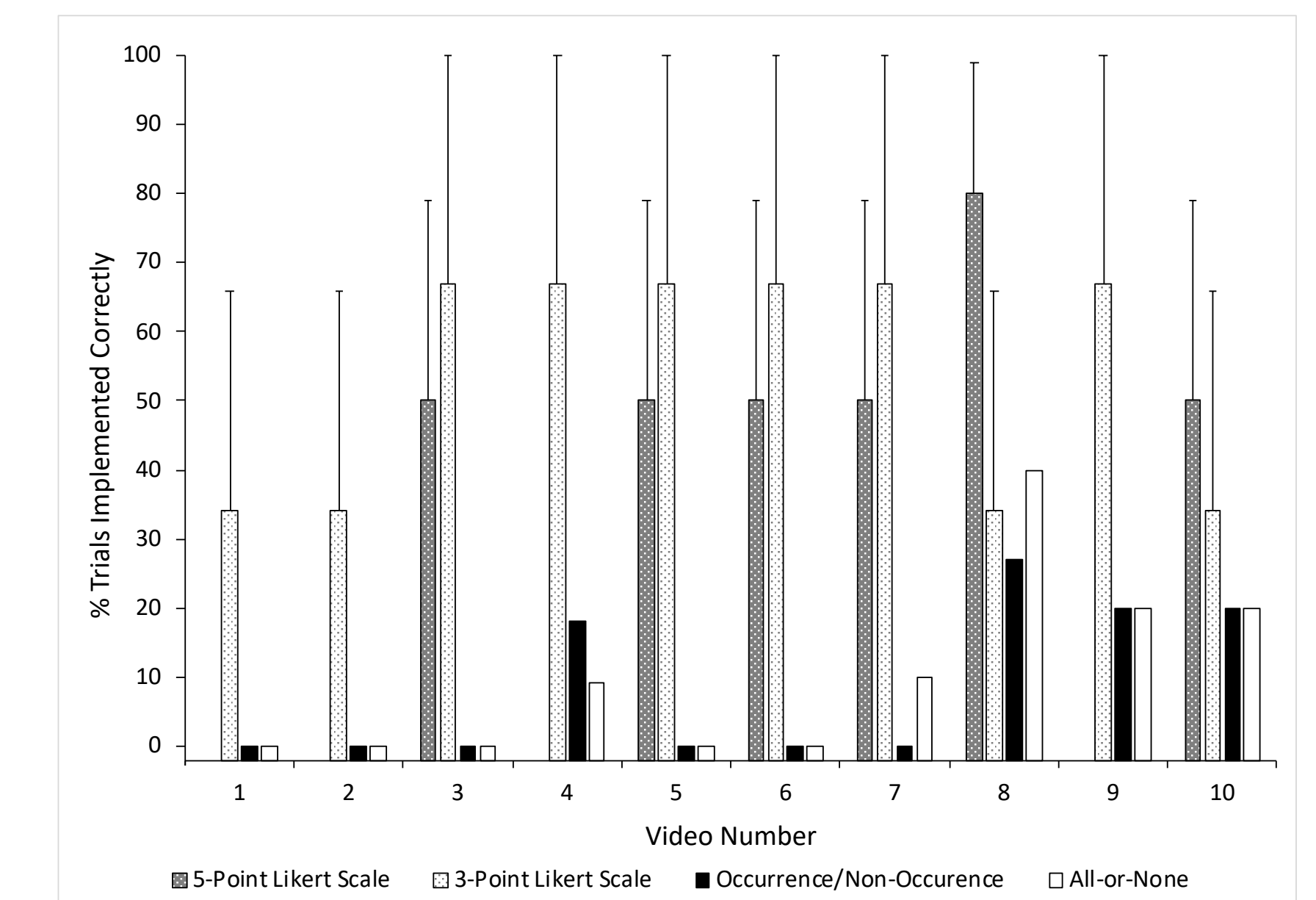


Figure 2. Percentage of trials implemented correctly across all scoring types for video samples 1-10. Overall Likert scores were assigned equivalent percentage ranges represented as error bars to evaluate agreement based on Suhrheinrich et al. (2019).

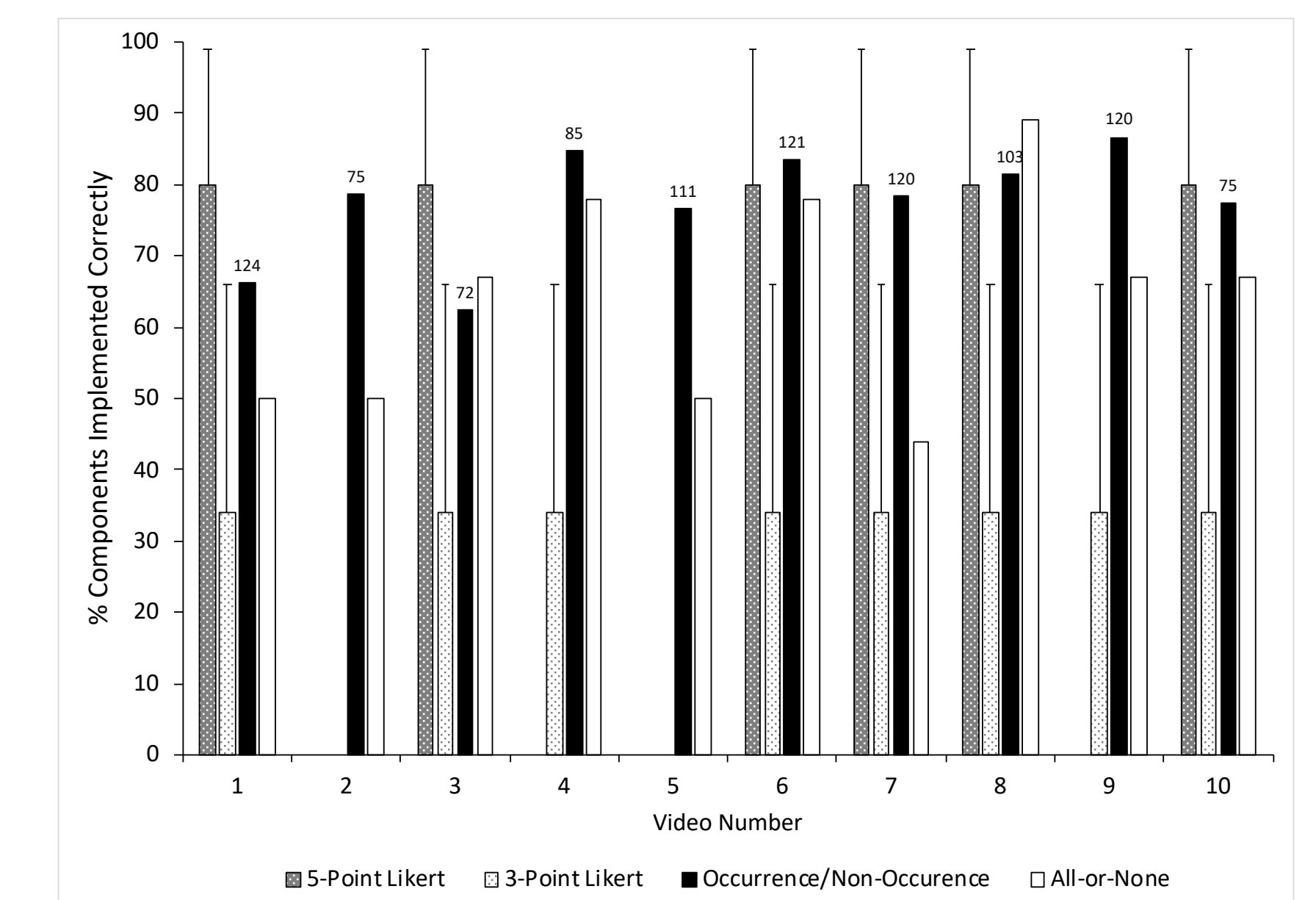


Figure 3. Percentage of components implemented correctly across all scoring types for video samples 1-10. Overall Likert scores were assigned equivalent percentage ranges represented as error bars to evaluate agreement based on Suhrheinrich et al. (2019).

## DISCUSSION

This project is currently ongoing. We aim to answer the following questions:

1. How specific is the information produced by each method?
2. Are some TI data collection methods more useful and efficient than others?
3. Can we get inter-rater reliability, and what is the overall reliability of each scoring type as compared to occurrence / non-occurrence?

## REFERENCES

- Carroll, R. A., Kodak, T., & Fisher, W. W. (2013). An evaluation of programmed treatment-integrity errors during discrete-trial instruction. *Journal of Applied Behavior Analysis, 46*(2), 379-394. 10.1002/jaba.49
- Gresham, F. M., MacMillan, D. L., Beebe-Frankenberger, M. E., & Bocian, K. M. (2000). Treatment integrity in learning disabilities intervention research: Do we really know how treatments are implemented? *Learning Disabilities Research, 15*(4), 198-205.
- Suhrheinrich, J., Dickson, K. S., Chan, N., Chan, J. C., Wang, T., & Stahmer, A. C. (2019). Fidelity assessment in community programs: An approach to validating simplified methodology. *Behavior Analysis in Practice, 13*, 29-39. <https://doi.org/10.1007/s40617-019-00337-6>