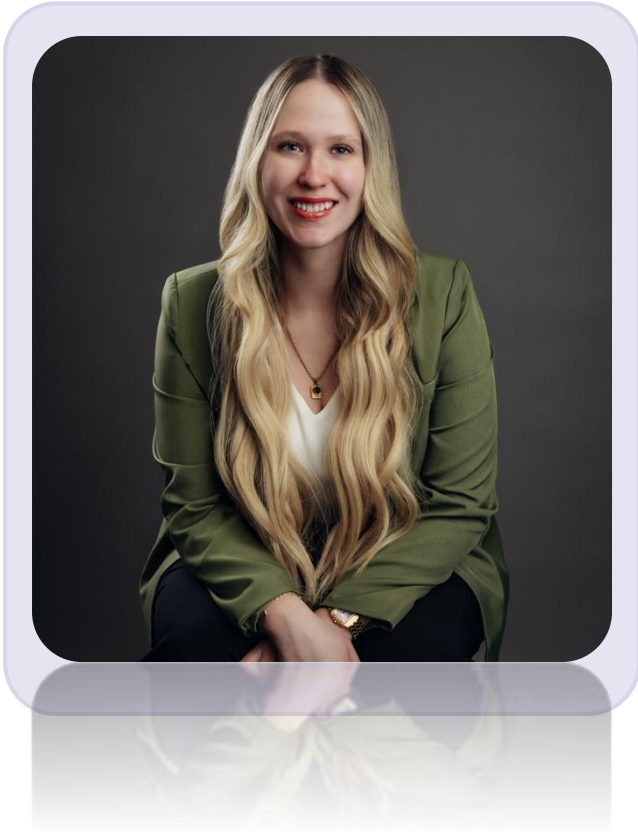




Treatment Integrity Is Not a Checkbox: Building Meaningful Fidelity Systems

Jessica Osos, PhD, BCBA-D, LBA

Affiliations & Disclosures



Jessica Osos, PhD, BCBA-D, LBA

- Doctoral level BCBA with research focused in generative learning strategies, toilet training, and clinical quality in ABA
- EIBI and ABA Clinical Program Manager for Intermountain Children's Health's ABA Services
- Independent consulting through Copper Consulting Group, LLC

Opinions are my own and do not represent any affiliated companies.

Upcoming CEUs

August 12th

Delegating with Confidence: Training and Supervising RBTs to Support Behavior Assessment

September 2nd

Beyond the Milestones: Selecting and Interpreting Curriculum-Based Assessments (VB-MAPP, ABLLS-R, and AFLS)

October 14th

TBD

November 12th

Crash Course in Behavior Analytic Toilet Training: Translating Research to Practice at AzABA Conference

Learning Objectives

1

Explain the functional relationship between treatment integrity, intervention effectiveness, and data-based decision making — including the ethical implications when fidelity is not directly assessed.

2

Identify common organizational, contextual, and behavioral variables that contribute to fidelity drift — including competing contingencies, training gaps, and system failures.

3

Describe and apply at least three empirically supported strategies for building sustainable treatment integrity systems: task analysis of procedures, direct fidelity measurement, and performance-based feedback.

Our Roadmap for Today

We will organize the hour around the practical questions a supervisor actually asks:

WHY

Why fidelity matters

WHO

Who collects the data

HOW

How it is measured

HOW OFTEN

How often it happens

HOW TO

How to do it well

BARRIERS

What gets in the way

01

Why Procedural Fidelity Matters

The case from theory, research, and ethics.

What Is Procedural Fidelity?

The definition

The extent to which an intervention is implemented as designed.

(Gresham, 1989; Vollmer et al., 2008)

- **A core component of internal validity.**

When behavior changes, fidelity data let us attribute that change to the intervention — not to confounds.

- **The basis for data-based decisions.**

High fidelity + no progress → revise the plan. Low fidelity → retrain the implementer. You cannot tell which without it.

- **An ethical obligation.**

Monitoring and evaluating our interventions is embedded across the BACB Ethics Code and CASP guidelines.

Why Measure Fidelity? Three Lines of Evidence

Descriptive

Paraprofessionals, teachers, and technicians make errors implementing interventions.

Carroll et al. (2013): DTI fidelity ranged from 21% to 74% in practice.

Correlational

As fidelity rises, client outcomes improve — and vice versa.

DiGennaro et al. (2007): students were off-task less when teachers implemented with higher fidelity.

Experimental

Programmed fidelity errors change how well and how fast procedures work.

Error type and degree affect efficacy across behavior-reduction and skill-acquisition procedures.

Two Kinds of Fidelity Errors

Errors of Omission

Failing to implement a prescribed component of the plan.

Example: the plan calls for a prompt on every trial, but the implementer skips it on some trials.

Carroll et al.: omission errors reliably slowed skill acquisition.

Errors of Commission

Adding a component that is not part of the prescribed plan.

Example: delivering reinforcement the protocol does not call for, or reprimanding when the plan does not include it.

St. Peter Pipkin et al.: commission errors were more detrimental when treating problem behavior.

Errors Have Idiosyncratic Effects

The same error can matter enormously for one client and barely at all for another — which is exactly why we measure.

21–74%

range of DTI fidelity observed
among paraprofessionals and
teachers (Carroll et al., 2013)

3

different acquisition outcomes for
3 clients under the same
commission-error condition
(Carroll et al.)

50%

fidelity — praising every other
correct response — already
counts as low fidelity for that
component

02

Where Practice Actually Stands

What two recent surveys of BCBAs reveal.

Two Recent Surveys, One Consistent Picture

203

Bergmann et al. (2024)

BCBAs supervising or providing ABA in any setting, any population

137

Colón et al. (2024)

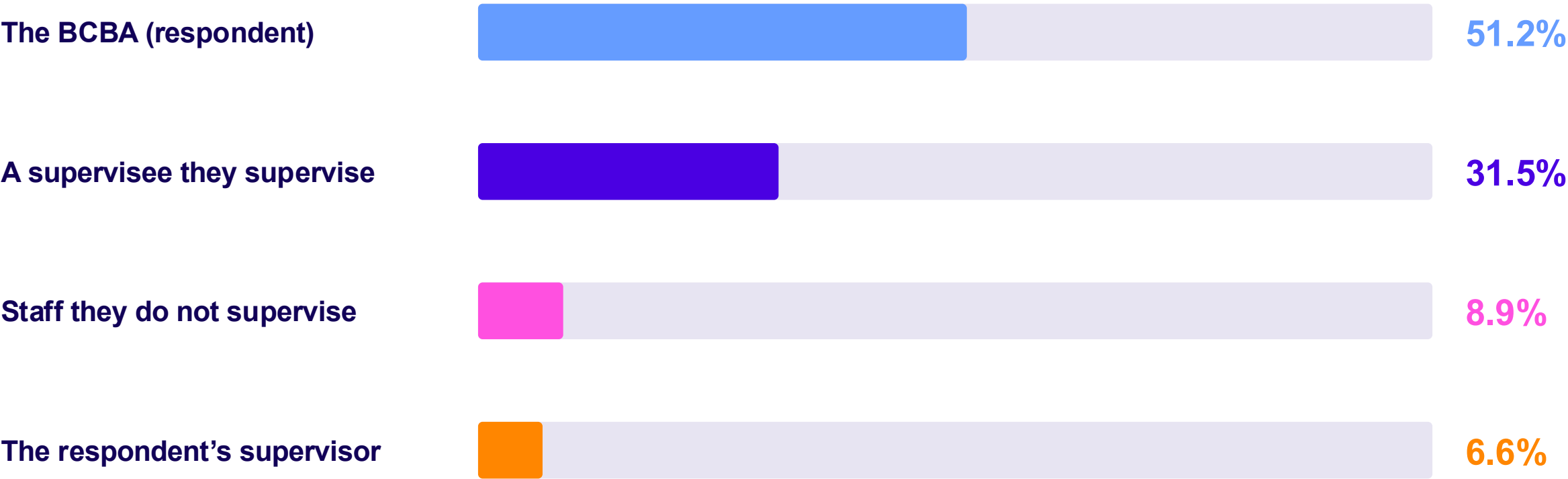
BCBAs providing ABA services, surveyed on training, practices, and barriers

Where they converge:

Nearly all respondents were trained in fidelity and value it — yet most collect data on only a small fraction of sessions, and the data-related steps (tracking, analysis) are the ones most often skipped.

Who Collects the Fidelity Data?

Bergmann et al. (2024), Table 3 — respondents could select more than one.



Takeaway: The BCBA and their direct supervisees do most of the collecting — training supervisees to collect fidelity data can expand coverage and build their repertoire.

How Is It Measured? Direct Observation Dominates

99.2%

use direct-assessment methods
to collect fidelity data

79.7% observe live — in person or
via streaming feed

Direct observation

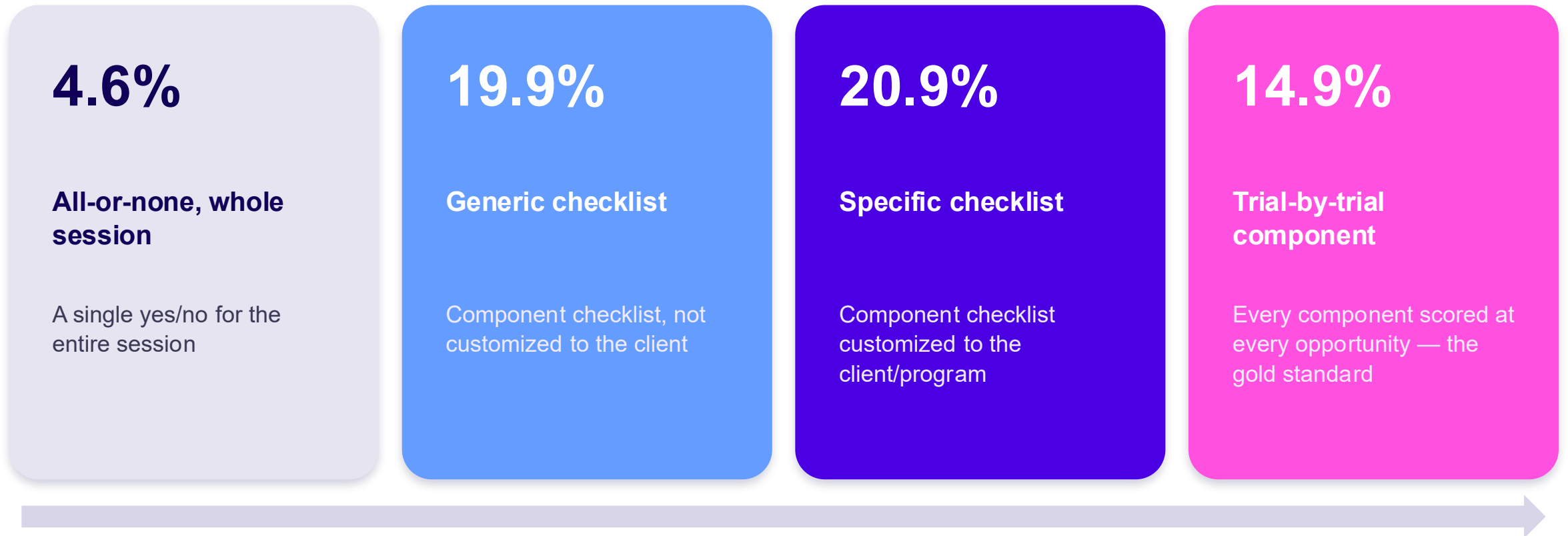
Watching the implementer deliver the procedure live or from a recording, and scoring against an operationalized checklist. It is the recommended method (Gresham et al., 2000).

The workhorse tool: the checklist

Combining specific, generic, and all-or-none component checklists, over half of respondents (55%) use a component-level dichotomous checklist — efficient and less likely to mask errors.

A Spectrum of Direct-Assessment Methods

From most global to most granular. More granular methods are more sensitive to incremental change — but more effortful.



more global / less effortful

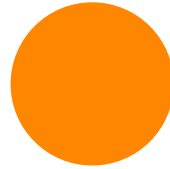


more granular / more sensitive to change

How Often? The 20% Problem

63%

of respondents collect direct-assessment fidelity data on 20% or fewer of their supervision sessions.



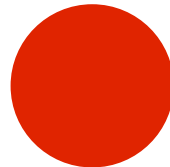
Only 8.4%

collect direct fidelity data for more than 80% of sessions.



~40% is the bar

Gresham et al. (2017) suggests fidelity be assessed for at least ~40% of sessions to yield a reliable estimate.



The gap

Few respondents reach even that threshold — so estimates may not reflect true fidelity.

Indirect Assessment: Useful, but Handle With Care

Only 30.1% use any indirect method. When they do, these are the tools — and the cautions.

39.7%

Permanent products

Inspect the environment after the fact (e.g., a completed data sheet).

Caution: Can only capture what leaves a lasting trace; may miss most components.

25.0%

Supervisee interviews

Ask the implementer how they ran the procedure.

Caution: May not improve fidelity; subject to recall and self-presentation.

22.1%

Supervisee self-reports

The implementer rates their own performance.

Caution: “Extremely reliable” but implementers tend to overestimate.

03

How to Measure It Well

A six-step, individualized framework.

A Six-Step Framework for Measuring Fidelity

Morris, Jones, & Oliveira (2024) — a practitioner's guide, built around direct observation.



Steps 1–2: Task-Analyze, Then Assign Measures

1

Task-analyze the procedure

Break the plan into individual, measurable components — sequential and conditional — each isolated so it can be scored on its own.

Operationalize fuzzy terms.

“Treat with compassion” isn’t measurable. Rescope it into observable behaviors (e.g., respectful salutations) so fidelity can actually be scored.

2

Assign a measure to each



Repeatability

count / frequency — e.g., praise delivered at least 5×



Temporal extent

duration — e.g., a 3–4 minute enriched break



Temporal locus

latency / IRT — e.g., redirect within 15 s

Global vs. Component Scores: Beware Masking

Scenario: A session has two components. Monitoring is perfect (100%), but praise is delivered on only 2 of 5 opportunities (40%).

Global measure

70%

Looks acceptable — and hides that a whole component was implemented at only 40% (Cook et al., 2015).

Component measure

Monitoring: 100% • Praise: 40%

Enables specific feedback: “you ran error correction consistently; let’s target praise.” That shapes performance far better than a single 70%.

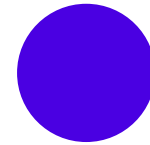
Step 3: Plan the Direct Observation

Direct observation is the gold standard — but it costs supervisor time. Design it to be efficient.



Prioritize

Monitor high-impact, complex components first — but sample simple ones too, because drift happens anywhere.



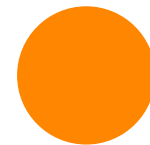
Choose the moment

Observe when the targeted components actually occur; rearrange schedules or record when needed.



Right-size length

Long enough for a representative sample of the target — a brief sample is often sufficient.



Set frequency

Use certification minimums (RBT 5%, etc.) as a floor, not a goal; balance low- and high-performers.

Step 4: Collect the Data — Mind Reactivity

The problem

Reactivity.

Staff may perform differently when a supervisor is watching, inflating the data and making it inaccurate (Kazdin, 1982).

Four strategies to reduce it (Reid et al., 2021):

- 1 Monitor frequently, so your presence isn't novel.
- 2 Begin observing immediately on entering the workspace.
- 3 Conduct unexpected "return" monitoring sessions.
- 4 Observe on a generally unpredictable schedule.

Steps 5–6: Analyze, Interpret, and Act

5

Analyze & interpret

Summarize into a global score for trend-tracking, but always retain component scores — global measures can obscure the errors that matter most.

Component scores drive better feedback:
reinforce what's strong, retrain only what's weak.

6

Take action



Low fidelity + no progress

Retrain the implementer — don't change a plan that was never tested.



High fidelity + no progress

Now it's appropriate to revise or replace the intervention.



Barrier surfaces in the data

Interview staff; adjust the plan to fit the real context (e.g., a 13-min interval).

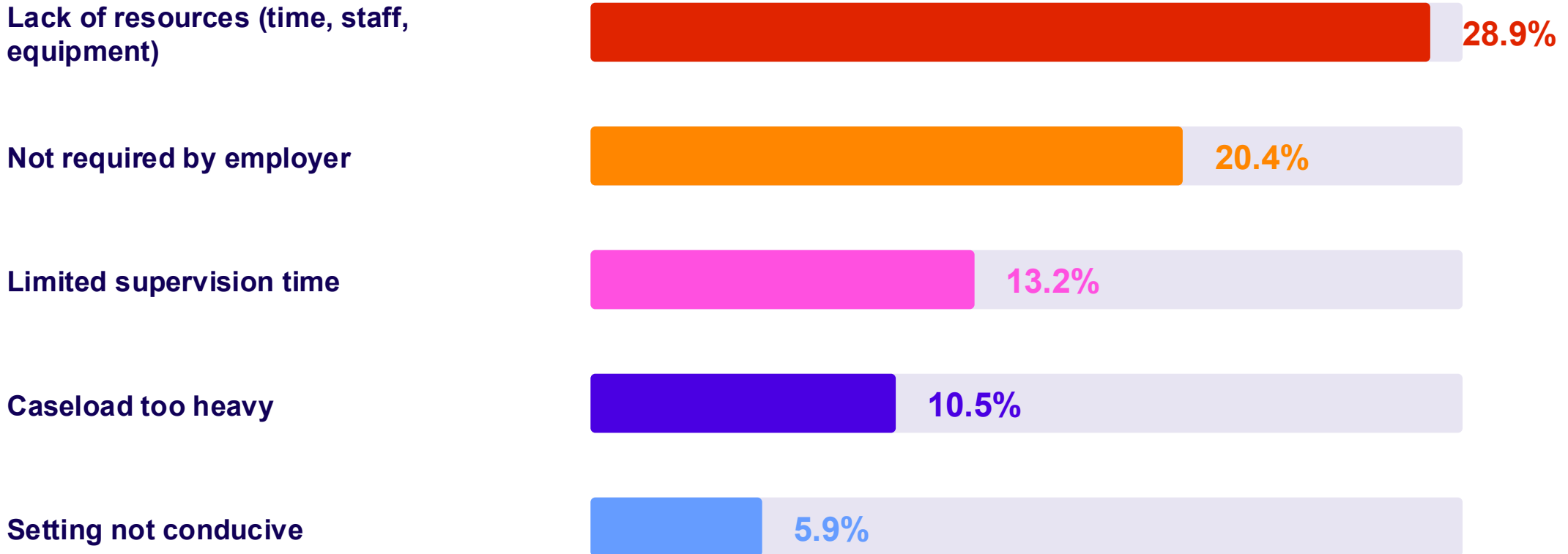
04

Barriers — and What to Do

Competing contingencies, training gaps, and fidelity drift — and how to fix them.

What Gets in the Way?

Barriers to assessing procedural fidelity in practice (Bergmann et al., 2024, Fig. 2; n = 152).



Echoed in Colón et al. (2024): the top barriers were no time, competing tasks taking precedence, and no established system — especially for data collection, tracking, and analysis.

The Training-to-Practice Gap

Well covered in training

Definition of fidelity 94%

Ethics code 95%

When/how to collect ~85%

Covered far less — and skipped most

How to track data 73%

How to analyze data 67%

How to provide feedback 58%

The pattern: the steps least covered in training — tracking, analysis, feedback — are the ones least likely to be done.

Closing the Gap: Practical Solutions



Build efficient tools

Use component checklists that capture the critical dimensions; consider systems that auto-graph data for later analysis.



Prefer specific checklists

Customized checklists detect errors better and enable personalized feedback — worth the setup time.



Train supervisees to collect

Expands coverage, builds an important task-list skill, and prepares future supervisors.



Set organizational expectations

Day-to-day expectations drive behavior. Make fidelity a required, resourced supervision activity (CASP).



Train with BST

Pair performance feedback with behavioral skills training to fix low fidelity, not just flag it.



Use resources that exist

Tutorials like Morris et al. (2024) provide ready templates and worked examples.

Pulling It All Together

Procedural fidelity is not a compliance checkbox — it is how we protect our clients and our conclusions.

- It underwrites internal validity, data-based decisions, and our ethical obligations.
- Most practitioners are trained and value it — yet collect it on too few sessions to be reliable.
- Direct observation with component-level checklists is the recommended approach.
- A repeatable six-step process makes individualized measurement feasible.
- The biggest barriers — time, systems, expectations — are organizational, and therefore solvable.

Reflection Questions

01

Which step — collection, tracking, or analysis — is the weakest link in your own practice right now?

02

Where in your organization could a shared, standardized fidelity system reduce everyone's effort?

03

What is one procedure you will task-analyze into a fidelity checklist this month?

Questions?

Thank you!

✉ **Email:** jess.osos@imail.org

📷 **Instagram:** [@jessicaosos](https://www.instagram.com/@jessicaosos)

🌐 **LinkedIn:** [/jessicaosos](https://www.linkedin.com/company/jessicaosos)