CHASE EXCELLENCE

Facilitator Scoring Guide

Reflective Practice Self-Assessment · Administration, Scoring, and Validity Notes

This guide is for facilitators and evaluators. Do not distribute it to participants: it contains the scoring criteria for the open-response items.

SECTION 1

What this instrument measures

The instrument has two parts and measures four constructs drawn directly from the logic model.

Items	Construct	What it captures
Part A, items 1–3	Movement I — Know Yourself	Self-awareness of one's own decision patterns
Part A, items 4–6	Movement II — Build Something	Distributing responsibility and preparing before reacting
Part A, items 7–9	Movement III — Build Beyond Yourself	Developing the capacity of others
Part A, items 10–12	Reflective Practice	The intended outcome: pausing and connecting decisions to principle
Part B, B1–B4	Measurement pathway	Recall, Principle Retrieval, Transfer, Behavioral Application

SECTION 2

Administration

- Administer at four points: baseline (before session one), end of series, and 90 and 180 days after the final session.
- Allow eight to ten minutes. Do not let participants consult the book or their notes for Part B.
- Participant codes let you match responses across time points without names. Collect the same code every time.
- For B3, read aloud or distribute one scenario from Section 5. Use a different scenario at each administration, and never use a scenario that appeared during the series.

CHASE EXCELLENCE | © 2026 Dr. Deangello Eley | chaseexcellencebook.com

- If a site is collecting supervisor corroboration of follow-through, obtain participant consent in writing before doing so.

CHASE EXCELLENCE | © 2026 Dr. Deangello Eley | chaseexcellencebook.com

SECTION 3

Scoring Part A

Each response scores 1 to 4: Rarely or never = 1, Sometimes = 2, Often = 3, Almost always = 4. No items are reverse-scored.

Compute a mean for each of the four subscales rather than a single total. Means preserve the 1–4 metric and stay interpretable if an item is skipped. Report subscale means separately; the Reflective Practice subscale (items 10–12) is the intended outcome and should be reported first.

Change is the difference in subscale mean between baseline and each later administration. Report it alongside the number of participants completing each administration, since attrition across four time points is the most common threat to interpreting these numbers.

Data quality check

Before analyzing, flag any form where all twelve responses are identical. Straight-lining indicates the participant did not engage with the items, and those forms should be reported separately rather than silently included.

SECTION 4

Scoring Part B

B1 is scored as a count: the number of sayings recalled without prompting. It needs no rubric.

B2, B3 and B4 are scored 0 to 3 using the rubrics below. Score only what the participant wrote. Do not credit what you believe they meant.

B2 — Principle Retrieval

Score	Criteria
0	No principle stated, or the response restates the saying without explaining it.
1	States a general idea loosely connected to the saying.
2	States the principle accurately, but only within the original story or context.
3	States the principle accurately and in general terms that extend beyond the original story.

CHASE EXCELLENCE | © 2026 Dr. Deangello Eley | chaseexcellencebook.com

B3 — Transfer

Score	Criteria
0	No saying selected, or the selection is unrelated to the situation.
1	Selects a plausible saying but cannot explain why it fits.
2	Selects an appropriate saying and explains why it fits the situation.
3	Selects an appropriate saying, explains the fit, and names a specific action it would prompt.

B4 — Behavioral Application

Score	Criteria
0	No specific decision described, or describes an intention rather than an action taken.
1	Describes a decision, but the connection to a saying or principle is unclear.
2	Describes a specific decision and names the saying or principle that informed it.
3	Describes a specific decision, names the principle, and states what would likely have happened otherwise.

Scoring reliability

Where results will be reported outside the building, have a second scorer independently score at least twenty percent of open responses and report the percentage of exact agreement. Discuss and resolve disagreements before finalizing scores.

CHASE EXCELLENCE | © 2026 Dr. Deangello Eley | chaseexcellencebook.com

SECTION 5

Transfer scenarios for B3

Use one scenario per administration. These situations are deliberately not drawn from the book, so that selecting a saying requires transfer rather than recall.

Scenario 1

A teacher on your team has asked you three times this month to sit in on conversations with a difficult parent. You are good at those conversations and they go better when you are there. The next request arrives on a day you are already behind.

Scenario 2

You are two weeks from a deadline on an initiative you proposed. A colleague points out a flaw you had not considered. Fixing it means telling your supervisor the timeline has to move.

Scenario 3

A staff member sends an email copying three other people, describing a decision you made in a way that is not quite accurate. You have a meeting with all of them in an hour.

A fourth administration requires a fourth scenario. Write one from an ordinary situation in your own setting, keeping it to three or four sentences and ensuring more than one saying could plausibly apply.

SECTION 6

Validity and reliability: what is and is not established

This section exists so that no one overstates what the instrument currently supports.

Established by design

- A documented content rationale. Every item was written to a specific construct in the logic model, and each construct is measured by three items. The mapping is shown in Section 1. This is a rationale, not evidence of content validity: establishing content validity requires independent expert review of item relevance and coverage, which has not been conducted.
- Face validity. Items describe observable behavior in plain language rather than attitudes or beliefs.
- Response format. A four-point frequency scale with no midpoint requires a directional judgment and reduces the tendency to select a neutral option. Frequency anchors are less vulnerable to acquiescence than agreement anchors.

CHASE EXCELLENCE | © 2026 Dr. Deangello Eley | chaseexcellencebook.com

Not yet established

- Content validity. Convene three to five people with relevant expertise who were not involved in writing the items. Ask each to rate every item for relevance to its stated construct and to identify anything the construct covers that no item captures. Revise accordingly and record what changed. Do this before the instrument is used for any reporting outside the building.
- Internal consistency. Once thirty or more completed baseline forms exist, compute Cronbach's alpha for each three-item subscale. Values at or above 0.70 support treating the subscale as a single measure; lower values indicate the items are not measuring one thing and should be revised.
- Item performance. Compute item-total correlations within each subscale. An item correlating below 0.30 with its own subscale is a candidate for rewriting.
- Construct validity. Establishing that the four subscales are distinct requires factor analysis on a substantially larger sample than a single site will produce.
- Sensitivity to change. Whether the instrument detects real change cannot be known until pre/post data exist across several cohorts.

Known limitations to state whenever results are reported

- Part A is self-report. Every item describes a behavior most educators would prefer to claim, so scores are subject to social desirability and are expected to be high at baseline.
- Part B partially offsets this, because B4 asks for a specific action and a counterfactual rather than a self-rating. Where the two disagree, treat Part B as the stronger evidence. B4 remains a proxy: it is a self-reported account of a past action, not an observation of one. An implementation commitment recorded during the series measures intention rather than behavior, and should never be reported as behavior.
- Neither part supports causal claims. Without a comparison group, change over time cannot be attributed to the book study, and results must be reported as change over time rather than as effects.
- The twelve Part A items are identical at every administration. Repeating an identical instrument can raise later scores on its own, independent of any change in practice. Say so whenever results are reported, and vary the open-response prompts if you administer more than twice.

© 2026 Dr. Deangello Eley. This resource may be reproduced and shared for noncommercial educational and professional-learning purposes with attribution.
It may not be sold, repackaged for sale, or represented as another person's work. Please retain the Chase Excellence branding and this notice. For the current version, visit chaseexcellencebook.com/resources.