

EPISODE 334

When To Trust A.I. As A School Leader, Part 4: Ratings



INSTRUCTIONAL
LEADERSHIP SHOW



Essential Question



To what extent can we trust A.I.
with teacher ratings?



Chef Evaluation: Grain Size



When evaluating a chef, would you evaluate:

- A dish?
- A meal?
- The chef's performance as an employee?
- The success of the restaurant?

Teacher Evaluation: Grain Size



When evaluating a teacher, would you evaluate:

- A walkthrough?
- A lesson?
- The teacher's performance as an employee?
- The success of the school?

Teacher Evaluation: Grain Size



When evaluating a teacher, would you evaluate:

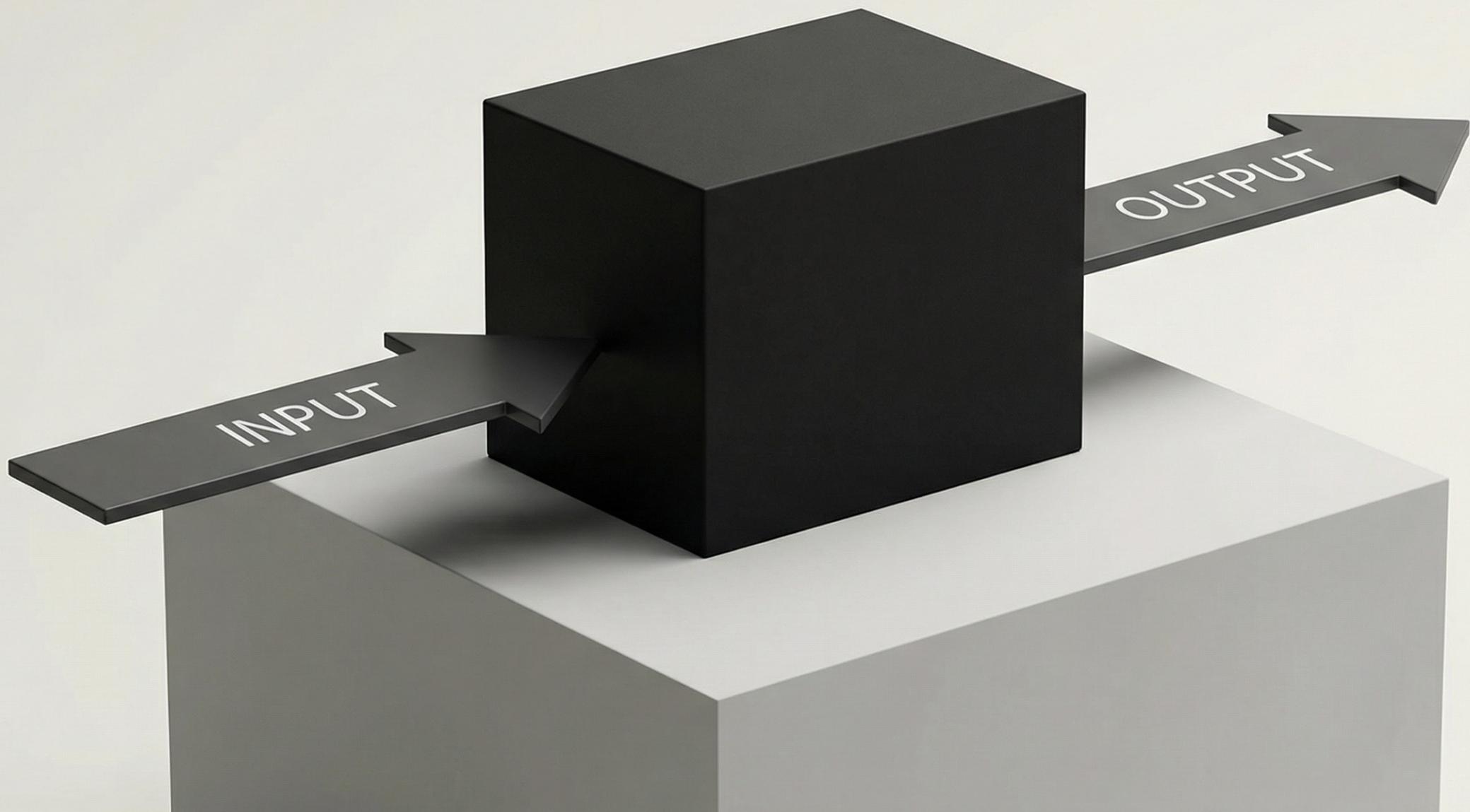
- A walkthrough?
- A lesson?
- The teacher's performance as an employee
- The success of the school?



Sampling in Teacher Evaluation



- We have required sampling procedures: formal observations and sometimes walkthroughs
- Other evidence sources, like lesson plans, goals and reflections, and student data are often explicitly admissible
- We may not officially be allowed to treat other impressions we've formed as evidence, but they shape our judgment
- Ultimately, the decision is a "black box" that we defend with evidence



Final Evaluations



- Satisfactory/Unsatisfactory or score (e.g. 1-4)
- Overall or by domain
- Determined by the evaluator's judgment
- Based on the totality of available evidence
- Some restrictions on admissible (citable) evidence

Qualitative \Leftrightarrow Quantitative



- Most evidence starts out as *qualitative* information
- We can apply tallies, percentages, and other fact-based calculations, manually or with AI reports
- We can apply subjective judgments to qualitative evidence to produce *quantitative* scores, e.g. 3 out of 4 on a rubric
- The number contains no new information, and is no more “true” than a qualitative judgment

A.I. Analysis



- Just one data point—take it with a grain of salt
- Check the facts (e.g. tallies & classification) on which it's based
- Stability is a concern—run multiple times

A.I. Reports & Scoring



Solutions ▾

Software ▾

AI ▾

Services ▾

Pricing

Login

Book Demo

Try Sibme

Closed vs. Open Ended Questions



Problem

Open-ended questions drive students towards deeper answers and demonstrate their critical thinking skills. However, it's not always clear who is responding to questions that promote more rigor. How can you efficiently measure and utilize this valuable aspect of student learning?



Deference Threshold



*The point at which we trust
a technology without
verifying its output.*