

## Socrates on Chicago's Failure

Scene: Hotel lobby

Characters: Socrates and Dr. Sidney Williams, influential education-policy-and-management entrepreneur

Socrates: Why is it that school districts are unable to institutionalize effective academic practices?

Dr. Williams: I think there is substantial data showing that school districts have made great strides in institutionalizing effective practices.

Socrates: Tell me more.

Dr. Williams: Well, the larger school districts are moving in the direction of doing more research and using evidence-based practices for teaching core content.

Socrates: Which specific districts are you referring to?

Dr. Williams: Chicago is representative of the trends in some larger cities. Chicago has a fairly large research and evaluation department, has initiated research, and has awarded contracts for evaluating key initiatives the district is pursuing.

Socrates: As I understand it the district is doing research on struggling readers.

Dr. Williams: True. But we refer to these students as “striving readers.” The district is in the process of conducting a large five-year study that incorporates everything we know to be effective.

Socrates: Like what?

Dr. Williams: Like smaller class size, more adults in the classroom, teachers networking and collaborating, parent involvement, increased

emphasis on in-service training, a corresponding increase in the training personnel—all bringing a strong focus on evidence-based practices.

Socrates: Well, it seems that the project covers many facets.

Dr. Williams: Yes, and the program has a particularly strong emphasis on emerging readers in grades 1-3, where the city uses a three-tier model to assure that children do not slip through the cracks. Each school has both a literacy impact teacher, and a lead literacy teacher.

Socrates: With all these strengths, it would seem that the program could not fail, unless, of course, the instruction is faulty. Tell me about the instructional programs that are used.

Dr. Williams: In all grades, teachers use high-quality, high-interest material. In all content areas teachers teach comprehension strategies. Teachers explicitly teach more vocabulary. A strong provision of the program is that lead literacy teachers are trained to use assessment and diagnostic data to adjust instructional practices. There are also after-school programs for students who need more help.

Socrates: I understand the research is conducted in quite a few schools.

Dr. Williams: A little over 30 experimental schools and 30 control schools.

Socrates: Has the city evaluated this model on a smaller scale before implementing it on such a grand scale?

Dr. Williams: Absolutely. The project is built around practices that have been shown to increase student performance. In fact, the city's chief instructional officer, Barbara Eason-Watkins, pioneered this effort when she was principal of an inner-city school, and she showed the impressive results that could be achieved.

Socrates: With such a large study and commitment to this all-encompassing approach, the district must feel quite sure that the results of the study will be highly positive.

Dr. Williams: Yes, that's why the study was designed so that it incorporated random assignment and other details needed for a pristine research study.

Socrates: Am I correct in assuming that the district's assumption is that if the Striving Reader initiative is successful, it will be institutionalized in the city?

Dr. Williams: Correct. The study has a potential to serve not only other Chicago schools but to be an institutionalized model for other cities that have serious problems with striving readers.

Socrates: Am I correct in assuming that the city would not institutionalize it if it is a failure?

Dr. Williams: Yes.

Socrates: Good. You indicated earlier that all instructional material was of high quality and high interest. What procedures are used to judge that they are of high quality?

Dr. Williams: One basis is the degree to which they are consistent with the adoption standards for high-quality programs. The other is the judgment of panels of knowledgeable teachers, administrators and others.

Socrates: Isn't this a rather circuitous way of judging the capacity of the program to affect student performance?

Dr. Williams: I fail to see that it is circuitous.

Socrates: Wouldn't it be far more direct to simply identify a program that has strong evidence of effectiveness, rather than infer the effectiveness from the degree to which it meets standards and endorsements of committees?

Dr. Williams: Well, that would be one way to do it, but certainly not the only way.

Socrates: What is the *best* evidence that a program is highly effective?

Dr. Williams: I don't know that there's any single best evidence.

Socrates: Couldn't the effectiveness of a program be completely determined by evidence that it is effective?

Dr. Williams: (Thinks.) I think that's probably the case... but...

Socrates: Could the effectiveness of a program be completely determined by information that it followed a particular formula for incorporating standards?

Dr. Williams: Yes, if the formula is a perfect predictor that following the formula always resulted in highly effective programs.

Socrates: But even if the formula is a perfect predictor, wouldn't you need actual results to confirm that the formula had been followed?

Dr. Williams: Yes, I suppose.

Socrates: So with the formula or without the formula, evidence of effectiveness is essential. Wouldn't it therefore be far more efficient to find programs that have been demonstrated to be effective and install them, rather than theorize whether a program will work and whether the formula you use is a perfect predictor?

Dr. Williams: Yes, if you were sure that the evidence was extensive enough. Possibly some important benefits of a particular approach are not captured by the data.

Socrates: Yes, that's a real possibility, but if the benefits are beyond whatever is measured, wouldn't it be possible to argue that any program has hidden benefits, regardless how poor its performance is on the things that are measured?

Dr. Williams: I suppose so, but...

Socrates: Here's another important question: would we possibly be aware of the program's unmeasured potential without some form of observable evidence?

Dr. Williams: No.

Socrates: If the evidence is observable, wouldn't we be able to measure it?

Dr. Williams: Probably, but possibly not.

Socrates: Then possibly you could describe a concrete example of something that has observable features that can't be measured.

Dr. Williams: I don't think I can come up with an example on the spot, but that doesn't mean that such possibilities don't exist.

Socrates: Exactly right. Nor does it mean that such possibilities do exist. But in a just world, what evidence do we use to judge that a program is *ineffective*?

Dr. Williams: Data on student performance.

Socrates: So if we know of student-performance data but don't know about other possibilities that may exist, which should we concern ourselves with?

Dr. Williams: It's not that simple. We see the instructional *process* as an end in itself. After all, this is what students will be engaged in. And the most beneficial process may be obscured if we simply look at outcome of effectiveness. Just as we would expect workers in a factory to produce evidence of effectiveness and measurable outcomes, we would not support a sweat shop that has inhumane provisions. For humanitarian health care issues and others are as important as output data.

Socrates: You make an excellent point. Certainly we should have rules that reflect humanitarian interests, but if the approach does not

break these rules, would we support health care and other issues if there were no evidence of effectiveness?

Dr. Williams: Probably not.

Socrates: So do we agree that the evidence of effectiveness is the *sine qua non* of instruction or manufacturing and that without evidence of effectiveness, the enterprise can't be endorsed as a success?

Dr. Williams: I don't think it's a simple yes-no question. It's a question of degree and direction.

Socrates: Could you present an example?

Dr. Williams: Yes, comprehension is a perfect example. A major focus of the Striving Readers program in Chicago is to make comprehension the core of *all* literacy instruction. There are seven areas of comprehension emphasis: summarizing, predicting, questioning, inferring, text structure, visualizing, and metacognition. It is important for students not only to learn these practices but to devote time applying and using them.

Socrates: Let's say we agree that all are reasonable endeavors. Would it follow that a particular emphasis should be pursued although there is no evidence of students learning it?

Dr. Williams: There may be disagreement about what constitutes learning it.

Socrates: If there is no evidence that students are learning a topic, wouldn't their performance suggest either that we drop the topic because it is apparently unteachable or fix the instruction so that students learn what they are scheduled to learn?

Dr. Williams: That may be a reasonable approach in some situations.

Socrates: Let's consider Chicago's situation with the Striving Reader Project: What kind of data does the program have so far?

Dr. Williams: The program has only been implemented two years, so there are still some growing pains, but there are very encouraging changes that set the stage for further growth.

Socrates: For example?

Dr. Williams: Well the project is being evaluated by an outside evaluator, Metis Associates, and their year-two report shows that many aspects of the model are being implemented. One very encouraging change is that the experimental classrooms almost universally have more adults in them than the control classrooms.

Socrates: But how has the model affected student performance?

Dr. Williams: It's probably too early to tell, but the first data question is whether the model is being implemented.

Socrates: Isn't it reasonable to assume that after two years of implementation, the project should show some positive impact on student performance? What is the student-performance data?

Dr. Williams: The impact of the program is not yet fully reflected by the student-performance data.

Socrates: Isn't it true that Metis' year-two evaluation report indicated that there was *no detectable difference* between the performance of experimental students and the control students?

Dr. Williams: I'm not sure that the statement was that strong, but it is true that the results are not as encouraging as the district had hoped.

Socrates: As I understand it, the first sentence of the executive summary conclusion was, "There are several possible reasons why an impact on student performance was not detected." Doesn't that indicate that there were no detectable differences between the control and experimental groups?

Dr. Williams: Yes, but as the sentence states there are possible reasons.

Socrates: Could one of those *possible* reasons be that the model is completely inept?

Dr. Williams: I suppose it's possible, but it doesn't seem very likely.

Socrates: Isn't it true that the Metis report strongly suggested that the model may be the cause of the poor data?

Dr. Williams: I haven't studied the report, so I don't know all the details.

Socrates: The details of this report may hold the key to what Chicago does next. It would seem that a plan B should be immediately implemented, particularly since no detectable improvement is observed after two years of Literacy Impact teachers, Lead literacy teachers, smaller class size, immersion in comprehension, and all the other variables that were supposed to accelerate students' performance. What is Chicago's plan B?

Dr. Williams: It's a little early to say, but I'm sure that some modifications are being considered.

Socrates: Do you perceive any sense of urgency in Chicago?

Dr. Williams: Certainly there's concern.

Socrates: Earlier, you indicated that the initiative would not be institutionalized unless it was effective. Given the current evidence of the program's apparent ineffectiveness, why wouldn't the next step be to direct the research department to discover which approach has been effective in Chicago and implement it, possibly as a comparison program in the Striving Reader study?

Dr. Williams: That seems reasonable, but I don't know how practical it is.

Socrates: Let me ask a more pointed question: If the city is committed to the principle of institutionalizing what works, why doesn't the city simply institutionalize Direct Instruction?

Dr. Williams: Oh, the current program is committed to direct instruction. There is extensive explicit teaching.

Socrates: I'm not referring to the district's conception of direct instruction, but the Direct Instruction programs Reading Mastery, Language for Learning, and Corrective Reading.

Dr. Williams: I don't understand why these particular programs would have priority over the current programs.

Socrates: Simply because they have a strong history of producing strong positive effects in Chicago.

Dr. Williams: Well, I know nothing about that.

Socrates: In the 60's, several Chicago schools had very good results using the field-tryout version of the program. Also, the Chicago Follow Through school, Ogden, scored in the top five of all Chicago schools. In the 70s, an entire district in Chicago used Direct Instruction and achieved superior results. In the 90s, the White Socks sponsored seven Chicago elementary schools that used DI, with highly positive results. Under Superintendent George Vallas, three demonstration schools achieved good results.

Dr. Williams: Now that you bring up that history, I recall that those Direct Instruction programs had been used in the city until fairly recently. I recall that they have been rejected.

Socrates: And what was the basis for this rejection?

Dr. Williams: Presumably data.

Socrates: As I understand it, the Chief Instructional officer Barbara Eason-Watkins, formed a committee to evaluate programs that were currently being used and vote on whether specific programs were to be eliminated. The committee voted against keeping Direct Instruction programs.

Dr. Williams: Well certainly they must have had evidence that it was not effective, which seems to contradict your assertion that Direct Instruction is effective. How do you explain this discrepancy?

Socrates: By referring to the Metis report. Nearly 2/3 of the data is on the degree to which the Striving Reader program is being implemented. Why would there be such a strong emphasis on documenting the implementation of the program?

Dr. Williams: Obviously, because if the program is not implemented well, the results would be misleading.

Socrates: Did the city document the degree to which Direct Instruction programs were being implemented before banning them?

Dr. Williams: I don't know.

Socrates: They didn't, nor did any of their predecessors who were opposed to DI. In fact, the sponsor of three demonstration schools during the Vallas era resigned from working in Chicago because the district was imposing inhumane procedures that prevented implementing the DI programs.

Dr. Williams: What were these so-called inhumane practices?

Socrates: One was that all reading instruction had to use grade-level material exclusively. Yet virtually none of the incoming children in the early grades performed at grade level. I believe that if you witnessed half an hour of this instruction and attended to the students' rate of failure, you would agree that it is inhumane.

Dr. Williams: Yes, that sounds unreasonable, but what does it prove about the current situation in Chicago?

Socrates: That there is apparent ignorance, and a transparent double standard. The city's current decision makers have no apparent knowledge that DI had been used very effectively in Chicago. This lack of memory is very convenient because it permits the district to experiment

with ideas that are the antithesis of DI practices. For programs the district wants to succeed, there are strong provisions for assuring that they are well implemented. For programs the district judges to be undesirable, there is a suspicious lack of concern with whether they are well implemented.

Dr. Williams: If you're suggesting that the district would purposely be unfair to a potentially successful program, I think you're going past the line of reason. We are dealing with administrators and technicians who are committed to doing what's in the students' best interest.

Socrates: Are you ruling out the possibility that Direct Instruction works?

Dr. Williams: Not at all. I'm simply suggesting that if it does work, its potential would be recognized.

Socrates: Consider the scales the district used to decide which approach had potential. On one side are more than a dozen DI schools in Chicago that achieved exemplary results. On the other side is one school. This single school totally outweighs those on the other side. Does that seem rational?

Dr. Williams: Eason's school didn't provide the only data used to make the judgment.

Socrates: Nor are the results of the Chicago DI schools the only data that DI has. Over 100 studies confirm the effectiveness of DI in improving reading performance. Here's a related question: if the district relied so heavily on the results of a single school, how closely was the data examined to assure that her achievement-test scores were valid?

Dr. Williams: What possible reason would anybody have for assuming that the data was not valid?

Socrates: It's called healthy skepticism. If so much hinges on the school's data being unassailable, a rational group would certainly put it

under the magnifying glass to document its effectiveness before committing over 30 schools to use it as a model.

Dr. Williams: Well I presume that the performance was scrutinized carefully.

Socrates: Possibly not. If the approach produced monumental gains in Eason's school but has had no detectable advantage in the Striving Reader study, isn't something seriously wrong with this picture?

Dr. Williams: The study is not completed yet.

Socrates: Are you saying that the apparent inconsistency between her school's achievement and that of more than 30 schools is not sufficient even now to raise a devil's-advocate question about the validity of her data?

Dr. Williams: The district may be investigating that possibility. I don't know.

Socrates: Here's a question that may help focus that possible investigation. Would that data be seriously suspect if it followed an impossible pattern?

Dr. Williams: Certainly. But I know of no such pattern.

Socrates: After she left the school, didn't the test scores drop perceptibly in the upper elementary grades?

Dr. Williams: I'm not sure. I know the school didn't do as well as it had done.

Socrates: A *very large* drop would signal that the earlier scores may have been bogus.

Dr. Williams: I don't see why, if the school was no longer using the same program she had installed. I could see that performance would deteriorate.

Socrates: If a fourth grade student performs at grade level and is taught nothing during the fifth-grade year, what performance would the student achieve on the fifth grade test?

Dr. Williams: The student would perform at the fourth grade level, and that's significantly behind.

Socrates: But if the student performed considerably below the fourth grade level—possibly the second-grade level or third-grade level—wouldn't that be highly suspicious?

Dr. Williams: Are you saying that happened?

Socrates: Check the records. Possibly the research department knows the answer.

Dr. Williams: Even if there was a severe performance drop, couldn't it be accounted for by high turnover of student population and low students entering the fifth grade?

Socrates: Certainly, but if you looked at the performance of continuing students, you could determine the extent to which those who had had high scores in the fourth grade had impossible scores in the fifth grade. And wouldn't such an investigation be very important?

Dr. Williams: Yes, but I'm not totally convinced that there's only one possible reason for such a performance drop.

Socrates: Indeed, one such possibility is that the new principal had methods that not only taught students nothing but erased skills and knowledge that were firmly implanted.

Dr. Williams: I thought we were having a serious discussion.

Socrates: I apologize, but my point is that the district manufactures many reasons for supporting favored initiatives; yet, for those that are not in favor, the slightest grey area becomes an excuse for disposing of them.

Dr. Williams: I think that's an exaggeration.

Socrates: But Chicago has committed thousands of children to participate in a study that wouldn't have a strong justification, even if the Eason-Watkins data were not suspect. Doesn't the district's behavior suggest that it is not really concerned about implementing what works; simply implementing what it wants to believe will work.

Dr. Williams: I return to the fact that your appeal ignores the fact that there is significant data beyond one school that supports details of this approach.

Socrates: Let's say that there is substantial data supporting Eason-Watkins's approach and substantial data supporting DI. Wouldn't it therefore be reasonable to run a fair test to see which approach is superior?

Dr. Williams: Yes, that would be reasonable.

Socrates: And wouldn't it be reasonable to pay the same amount of attention to implementation of both programs, so that one doesn't gain the advantage of being implemented far better than the other?

Dr. Williams: Mmmm. Yes.

Socrates: Why isn't that design being considered now?

Dr. Williams: Again, you're making it sound as if that's the only rational possibility, when there are many programs other than Direct Instruction programs that claim data of effectiveness.

Socrates: An excellent point. Then why not test them also and get facts on what actually works best in Chicago?

Dr. Williams: The district may be considering this possibility. I can't speak to its agenda.

Socrates: It would seem that its agenda should be urgent.

Dr. Williams: I agree.

Socrates: One other issue. I would have judged that a knowledgeable superintendent would have insisted on a stronger research

design, like the one that tests different possible approaches rather than testing one approach against baseline. Who was superintendent of Chicago when this research plan was formulated?

Dr. Williams: Arne Duncan, and his proper title was CEO, not superintendent.

Socrates: My mistake. And what's Arne Duncan's proper title now?

Dr. Williams: Secretary of Education for the United States.

Socrates: And who did he nominate to succeed him as CEO of Chicago public schools?

Dr. Williams: Barbara Eason-Watkins.

Socrates: So he seems to be a believer in the Eason-Watkins philosophy.

Dr. Williams: Definitely, but are you implying that the two were in collusion?

Socrates: No. Possibly one of them supports "radical" test preparation and the other is simply naïve.

End