



## **TRANSCRIPT**

**Rand Corporation's Leading  
Education Research**

**Dr. Stephen Klein Raises  
Grave Doubts About Texas  
Testing Integrity**

**National Academy of  
Sciences Presentation**

**June, 1999**

**Irvine, California**

## Reporter's Certificate

\* \* \* \* \*

I, Beth C. Drain hereby certify:

that on the 2nd & 3rd day of  
June, 1999, I did report in  
shorthand the testimony of the foregoing  
proceedings;

that on the conclusion of the above entitled  
matter, I did transcribe my shorthand notes into  
typewriting;

that the foregoing transcript is a true and correct  
record of my shorthand notes thereof.

Beth C. Drain

Certified Shorthand Reporter

Certificate No. 7152

527  
481



1 VERY HIGH FOR THEM TO BE ABLE TO QUALIFY FOR THIS  
2 AND BE ABLE TO KEEP THE WAIVER. AND FROM YEAR TO  
3 YEAR THE COMMISSIONER REVIEWS THE WAIVER TO SEE IF  
4 THE'RE MOVING THAT WAY. IF THEY HAVEN'T MET IT,  
5 THE COMMISSIONER HAS THE OPTION OF PULLING THE  
6 WAIVER, AND THEY GO BACK TO HAVING TO GO TO A  
7 TARGETED ASSISTANCE MODEL. THANK YOU.

8 MR. NEISSER: THANK YOU. NOW WE'RE GOING  
9 TO WAIVE THE NEXT COUPLE OF SECTIONS AND MOVE --  
10 YOU'LL BE HAPPY TO HEAR -- TO WHAT IS CALLED  
11 BREAK. IT'S SUPPOSED TO HAPPEN AT 3:15 AND IT'S  
12 ONLY A LITTLE LATER THAN THAT NOW. AND WE'LL  
13 RESUME IN ABOUT FIVE MINUTES WITH REACTIONS,  
14 STEPHEN KLEIN AND JOE JOHNSON AND MARTIN CARNOY.

15 (RECESS TAKEN.)

16 MR. NEISSER: THANK YOU, FOLKS. THANK YOU  
17 FOR COMING BACK IN SO PROMPTLY. OUR NEXT IS TO  
18 HAVE THREE COMMENTATORS ON THE TEXAS PROGRAM, AND  
19 THE FIRST OF THE COMMENTATORS IS STEVE KLEIN OF  
20 RAND.

21 MR. KLEIN: MONTY PYTHON, AND NOW YOU GET  
22 SOMETHING COMPLETELY DIFFERENT. TO GIVE YOU  
23 SOMETHING A LITTLE DIFFERENT. RAND IS CONDUCTING A  
24 PROJECT RIGHT NOW AT ELEVEN SITES AROUND THE  
25 COUNTRY. ONE OF THOSE SITES -- SOME OF THESE SITES



1 ARE SCHOOL DISTRICTS, SOME OF THEM ARE WHOLE  
2 STATES. ONE OF THOSE SITES IS IN TEXAS, WHICH IS  
3 WHAT BROUGHT ME HERE TODAY.

4 WHAT I'M GOING TO TALK ABOUT TODAY IS  
5 NOT RELATED TO THE PROJECT DIRECTLY. IT WAS SORT  
6 OF A SERENDIPITOUS FIND. THE PROJECT HAS TO DO  
7 WITH THE RELATIONSHIP BETWEEN STUDENT PERFORMANCE  
8 AND THE USE OF REFORM PRACTICES IN MATH AND  
9 SCIENCE. SO WE WERE TESTING, AS YOU KNOW, THERE'S  
10 A BIG PROGRAM OUT THERE NATIONAL SCIENCE FOUNDATION  
11 HAS SPENT ABOUT A BILLION DOLLARS WHICH IS BASED  
12 UPON THE PREMISE THAT IF TEACHERS ENGAGE IN REFORM  
13 PRACTICES, THEIR STUDENTS WILL DO BETTER IN MATH  
14 AND SCIENCE. NOBODY EVER CHECKED THAT ASSUMPTION.  
15 SO WE ARE RUNNING A PROJECT THAT IS INVESTIGATING  
16 THAT. AND PART OF THAT PROJECT, LITTLE SMALL PIECE  
17 OF THAT PROJECT, INVOLVED ADMINISTERING OPEN-ENDED  
18 AND MULTIPLE-CHOICE TESTS TO STUDENTS.

19 AND WHEN WE WENT TO DO THAT, WE WERE  
20 SOMETIMES TOLD THAT, NO, YOU COULDN'T GO INTO THE  
21 SCHOOLS DURING THIS PERIOD OR THAT PERIOD OR SO ON.  
22 AND ON ONE OF THOSE OCCASIONS, WE ASKED -- WE CAN'T  
23 GO IN THERE, IS THAT BECAUSE THEY'RE DOING  
24 STATEWIDE TESTING? AND THE ANSWER WAS NO. THAT'S  
25 WHEN THEY'RE PREPARING TO DO THE STATEWIDE TESTING.



1 THEY HAD BROUGHT IN PRIVATE CONTRACTORS.

2 AND I SUBSEQUENTLY MET QUITE RECENTLY  
3 WITH SOME OF THOSE CONTRACTORS. THEY'RE THE SAME  
4 PEOPLE WHO TEACH KIDS HOW TO TAKE THE COLLEGE BOARD  
5 EXAMS. WE'RE GOING TO TELL YOU HOW TO BEAT THE  
6 TEST.

7 WELL, GETTING BACK TO MY ORIGINAL  
8 PART, IN DOING THE STUDY, WE WERE VERY SENSITIVE TO  
9 THE FACT THAT PEOPLE WERE CONCERNED THAT  
10 MULTIPLE-CHOICE TESTS JUST SEEMED TO MEASURE  
11 ECONOMIC ADVANTAGE. THERE'S THIS STRONG  
12 RELATIONSHIP BETWEEN STUDENT PERFORMANCE AND  
13 INDICATORS OF SOCIOECONOMIC STATUS, SUCH AS THE  
14 PERCENTAGE OF KIDS ON FREE AND REDUCED LUNCH. JUST  
15 THINK ABOUT THE BIG FLAP OVER THE NCAA PUTTING IN  
16 SAT SCORE REQUIREMENTS. WHAT WAS THE MAJOR  
17 CONCERN? IT'S BECAUSE OF THIS CLOSE RELATIONSHIP  
18 BETWEEN ECONOMIC ADVANTAGE AND TEST SCORES.

19 AND EVEN IN SOME OF THE MATERIALS  
20 THAT YOU HAVE, THERE'S NEWSPAPER ARTICLES AND  
21 PICTURES AND GRAPHS OF THIS VERY STRONG  
22 RELATIONSHIP. AND I'LL GIVE YOU SOME EXAMPLES OF  
23 THAT.

24 HERE IS -- EACH OF THESE DOTS  
25 REPRESENTS A SCHOOL OR CAMPUS, TERMS HAVE BEEN USED



HERE. ALONG THE -- SORT OF SEE THE TREND LINE THERE. PRETTY STRONG. NO BIG SURPRISE HERE, RIGHT? HERE IS THE MEAN SCIENCE SCORE. THIS IS A MULTIPLE -- THIS IS A HANDS-ON TEST THAT WAS DEVELOPED IN CONJUNCTION WITH STANFORD AND RAND, RICH SHABLESON AND BRIAN AND A NUMBER OF OTHER PEOPLE WORKED ON IT, VERY OPEN-ENDED KIND OF THING. KIDS DO EXPERIMENTS AND THEY RECORD THE RESULTS IN A BOOK AND SO ON, PUSHING TOWARDS THE REFORM PRACTICES. EVEN WITH THAT KIND OF TEST, WE SEE THIS STRONG RELATIONSHIP BETWEEN TEST SCORES OVER HERE AND THE PERCENT OF STUDENTS ON FREE AND REDUCED LUNCH AT THE SCHOOL.

THIS SHOULD COME AS NO SURPRISE. EVERYBODY, THIS IS EXACTLY WHAT YOU WOULD NORMALLY SEE, AND THAT'S WHAT WE FOUND.

AND WE GET THE SAME STORY WHEN WE TURN TO A TEST THAT WAS -- SO THE PICTURE, FREE AND REDUCED LUNCH, SCIENCE -- BUT THIS IS A MULTIPLE-CHOICE TEST. AND THE CORRELATION IS A LITTLE BIT STRONGER IN PART BECAUSE IT'S A MORE RELIABLE TEST THAN THE OPEN-ENDED HANDS-ON THING. WE CAN TALK ABOUT THAT AT SOME POINT. BUT THE RELATIONSHIP IS PRETTY MUCH WHAT YOU SAW, AND THE OTHER ONE BETWEEN THE FREE AND REDUCED LUNCH AND



1 TEST SCORES, THEN WE SEE THIS NEGATIVE CORRELATION.  
2 AND BOTH OF THESE TESTS WERE ADMINISTERED BY PEOPLE  
3 WE HAD HIRED TO GO INTO THE SCHOOLS TO DO THIS.

4 AND HERE'S THE SAME STORY. THE FIRST  
5 TWO GRAPHS WERE IN MATH. HERE'S IS THE SAME STORY  
6 BASICALLY. EXCUSE ME. THE FIRST TWO WERE IN  
7 SCIENCE, AND HERE'S THE SAME STORY IN MATHEMATICS.  
8 RELATIVELY STRONG NEGATIVE CORRELATION BETWEEN TEST  
9 SCORES AND PERCENTAGES OF STUDENTS ON FREE AND  
10 REDUCED LUNCH. NO SURPRISES HERE.

11 THEN WE TOOK A LOOK AT THE  
12 CORRELATION BETWEEN THOSE SAME MEASURES AND THE  
13 TAAS, AND IT BLEW UP. LOOKS LIKE SOMEBODY HAD HIT  
14 THIS THING WITH A SHOTGUN. FREE AND REDUCED LUNCH.  
15 HERE IS MEAN MATH, TAAS SCORES, AND THE CORRELATION  
16 IS A .04. NO RELATIONSHIP AT ALL. SAME KIDS.  
17 EXACTLY THE SAME KIDS TWO WEEKS LATER. EXACTLY THE  
18 SAME KIDS TWO WEEKS LATER, WE HAD INDIVIDUAL  
19 SCORES.

20 → HERE'S ANOTHER PICTURE. THIS IS  
21 READING SCORES IN THESE SAME SCHOOL DISTRICTS --  
22 SAME SCHOOLS RATHER. SAME SCHOOLS. AND WHAT IS  
23 INTERESTING ABOUT THIS ONE, IT SEEMS LIKE WE HAVE  
24 TWO THINGS GOING ON. WE HAVE THIS TREND HERE, AND  
25 THEN WE HAVE THIS CLUSTER UP HERE. THAT'S ACTUALLY



TEN SCHOOLS. EIGHT DOTS, BUT TEN SCHOOLS THERE BECAUSE TWO OF THE SCHOOLS FALL IN EXACTLY THE SAME PLACE AS TWO OTHER SCHOOLS. YOU TAKE THOSE OUT, CORRELATION JUMPS RIGHT BACK UP. YOU TAKE OUT THAT CLUSTER OF SCHOOLS, EVERYTHING CHANGES AND GOES BACK UP TO A MINUS 91, WHICH IS WHERE YOU WOULD EXPECT IT TO BE.

NOW, THIS IS NOT AN OUTCOME THAT WE WANTED TO FIND AT ALL BECAUSE THIS POSES REAL PROBLEMS FOR US BECAUSE WE HAD HOPED TO USE THE TAAS SCORES. I DON'T FEEL COMFORTABLE DOING THAT ANYMORE GIVEN THESE RESULTS BECAUSE I THINK THE SCORES ARE SUSPECT. THERE'S LOTS OF POSSIBLE EXPLANATIONS FOR WHAT HAPPENED, NOT ONE, BUT MANY POSSIBLE EXPLANATIONS FOR WHAT OCCURRED HERE.

BUT WHAT GOT ME THINKING ABOUT THIS SITUATION IS HOW DIFFERENT IT IS THE WAY WE GO ABOUT TESTING KIDS IN SCHOOLS FROM THE WAY WE GO ABOUT GIVING TESTS THAT REALLY MATTER, LIKE THE KIND OF WORK THAT I DO IN LICENSING LAWYERS OR DOCTORS OR CPA'S OR PILOTS OR OTHER KINDS OF THINGS. WHEN I WORK ON A LICENSING TEST, IT'S TOTALLY DIFFERENT THAN THE KINDS OF TESTS WE USE IN SCHOOLS. WE HAVE A THIRD PARTY ADMINISTER THE TEST.





1 SAM -- THE LATE SAMPLE MESSICK WAS AT  
2 A PANEL WITH JAY AND I -- AT THIS TIME OF YEAR --  
3 SITTING THERE AND SAM SAID THIS IS THE FOX GUARDING  
4 THE HEN HOUSE, HAVING TEACHERS ADMINISTER THE  
5 TESTS. IT IS TOTALLY DIFFERENT. BY THE WAY, WE  
6 FOUND THE SAME RESULTS IN LOUISIANA, BUT WE DIDN'T  
7 FIND THEM IN OHIO. SO I'M NOT SAYING THAT THIS IS  
8 UNIVERSAL. I'M NOT SAYING THE PROBLEM IS IN EVERY  
9 SCHOOL OR IN EVERY TEACHER. BUT THERE'S THINGS  
10 WHICH ARE GOING ON OUT THERE WHICH ARE PRODUCING  
11 THAT .04 CORRELATION WITH THE SAME KIDS, SAME  
12 TESTS, SAME KINDS OF TEST, DIFFERENT FORM.

13 NOW, WHAT COULD BE GOING ON? WELL,  
14 IT COULD BE THE BREAK IN THE STANDARDIZATION OF THE  
15 TEST. TIME LIMITS COULD FLEX. - TEACHER COULD WALK  
16 AROUND THE ROOM POINTING AT THE PAGE. TEACHER  
17 COULD BE DOING THINGS LIKE GIVING STUDENTS EXAMPLE  
18 PROBLEMS THAT ARE VERY SIMILAR TO THE KINDS OF  
19 PROBLEMS THEY'RE GOING TO SEE ON THE TEST.

20 RECENTLY THERE WAS A BIG FLAP IN  
21 RHODE ISLAND ABOUT TEACHERS TEACHING THE ACTUAL  
22 TEST. AND THE SOLUTION WAS WHAT WE'RE GOING TO DO  
23 IS WE'RE GOING TO PUT A BAR CODE ON EVERY TEST  
24 BOOKLET, THAT SOMEHOW THIS IS GOING TO SOLVE THE  
25 PROBLEM. APPARENTLY THE PEOPLE IN RHODE ISLAND



HADN'T HEARD ABOUT XEROX MACHINES. I CAN'T QUITE  
FIGURE THAT ONE OUT.

BUT WE WERE JOKING DURING LUNCH ABOUT  
ALL THINGS THAT WE DEAL WITH WITH TEST SECURITY ON  
HIGH-STAKES TESTS LIKE COLLEGE BOARD EXAMS OR THE  
LAW SCHOOL ATTITUDE TEST OR LICENSING EXAMS AND HOW  
DIFFERENT THE WHOLE PHILOSOPHY IS. CHANGE THE TEST  
FORMS, WE EMBED THINGS INTO THE TEST THAT WE CAN  
CHECK FOR UNAUTHORIZED BEHAVIOR. WE CAN DO ALL  
SORTS OF THINGS ON THESE LICENSING TESTS THAT WE  
DON'T DO AT ALL.

CALIFORNIA GAVE A TEST LAST YEAR, THE  
STANFORD IX, TO THREE MILLION KIDS IN CALIFORNIA.  
I HAVE BET THAT THE SCORES WILL BE HIGHER. THEY'RE  
GOING TO GIVE THE SAME TEST OR THEY GAVE THE SAME  
TEST. I HAVE BET MY CONDO THAT THE SCORES WILL BE  
HIGHER. NOBODY IS GOING TO TAKE ME UP ON THAT, ON  
THAT BET. THE SCORES WILL BE HIGHER. IT'S THE  
SAME TEST. IS THERE ANYBODY WHO DOUBTS THAT THE  
SCORES ARE GOING TO BE HIGHER? THE GOVERNOR IS  
GOING TO BE -- THE NEW GOVERNOR IS GOING TO BE  
REALLY HAPPY. LOOK IT. I CAME IN AND LESS THAN A  
YEAR THE SCORES WENT UP DRAMATICALLY. THIS HAPPENS  
EVERY OTHER STATE.

I THINK THAT WHEN WE LOOK AT THE --



1 BY THE WAY, THE MORE PRESSURE YOU PUT ON THE  
2 ACCOUNTABILITY SYSTEM, THE BETTER THE REWARDS, THE  
3 MORE IMPORTANT THEY ARE, THE MORE IT'S PUBLICIZED,  
4 THE MORE YOU DO THIS, THE MORE YOU ARE GOING TO BE  
5 PUSHING SOME TEACHERS, SOME PRINCIPALS, SOME  
6 SCHOOLS TO ENGAGE IN BEHAVIOR THAT THEY THINK IS  
7 QUITE OKAY.

8 ONE TEACHER SAID TO ME, QUITE  
9 FRANKLY, WELL, OF COURSE, WE'RE TEACHING THE KIDS  
10 THIS STUFF. THIS IS WHAT THEY'RE GOING TO BE  
11 TESTED OVER. WHY WOULDN'T I TEACH THEM THAT?

12 GOING UP TO ARA IN MONTREAL, SO MANY  
13 OF YOU WERE THERE, I WAS FLYING UP FROM BOSTON AND  
14 I WAS READING THE AIR CANADA MAGAZINE. AND IN THE  
15 AIR CANADA MAGAZINE THEY HAD FACTS ABOUT AIR  
16 CANADA'S FLYING INTO TORONTO, AND THEY NOTED THAT  
17 EVERY HOUR -- ON THE AVERAGE EVERY HOUR THERE WERE  
18 FIVE PLANES, FIVE AIR CANADA FLIGHTS ARRIVING AT  
19 TORONTO INTERNATIONAL AIRPORT. THE NEXT ITEM ON  
20 THE LIST OF THESE FACTS WAS THAT ON THE AVERAGE  
21 EVERY HOUR, FOUR AIR CANADA FLIGHTS WERE TAKING  
22 OFF. THIS IS TRUE. REALLY HAPPENED.

23 AND I WAS THINKING ABOUT THIS. THE  
24 POSSIBLE EXPLANATION, BY THE WAY, WAS THE VERY LAST  
25 ITEM ON THE LIST OF FACTS WAS THAT THERE WAS 600



1 TAXICABS HAILED A DAY AT TORONTO, PLANES LANDING,  
2 ONE OUT OF FIVE, THE WINGS WOULD BE CUT OFF AND  
3 CONVERTED TO A TAXI AND TAKE OFF.

4 BUT I WAS THINKING ABOUT THIS WHILE  
5 BILL SCHMIDT WAS TALKING ABOUT TENS MATH. BILL WAS  
6 POINTING OUT THAT, YES, THE US WAS BEHIND MANY  
7 OTHER COUNTRIES, BUT IT WASN'T SO BAD IN SOME  
8 AREAS, IT WAS WORSE THAN SOME AREAS, BETTER THAN  
9 SOME AREAS, PARTICULARLY GOOD IN ROUNDING, WHICH  
10 MAYBE CANADA SHOULD THINK ABOUT. THE NOTION IS,  
11 YES, YOU CAN PROBABLY TEACH STUDENTS HOW TO TAKE A  
12 PARTICULAR TEST OR A PARTICULAR TYPE OF TEST. IT'S  
13 NOT NECESSARILY SAYING THEY'RE TEACHING EXACTLY THE  
14 SAME QUESTIONS, BUT MODEL QUESTIONS THAT PARALLEL  
15 VERY CLOSELY THE KINDS OF QUESTIONS THEY'RE GOING  
16 TO SEE ON THE TEST, AND YOU CAN LIMIT INSTRUCTION  
17 TO THAT.

18 AND TO THE EXTENT THAT THAT OCCURS,  
19 IT OCCURS IN SOME SCHOOLS MORE THAN OTHERS. THE  
20 INTERESTING THING ABOUT THIS CHART TO ME WAS IT WAS  
21 OCCURRING -- WHAT APPEARS TO BE THE INAPPROPRIATE  
22 BEHAVIOR WAS OCCURRING AT THE SCHOOLS WHICH ARE  
23 LIKELY TO HAVE THE LARGEST CONCENTRATION OF  
24 MINORITY STUDENTS. THE SCHOOLS THAT ARE GOING BE  
25 REALLY UNDER THE GUN TO PUSH SCORES UP ARE THE ONES



1 THAT ARE GOING TO BE TOWARD THE BOTTOM END OF THE  
2 SCORE DISTRIBUTION.

3 SO ALL I'M DOING IS SAYING BE CAREFUL  
4 OUT THERE FOR THIS PUSH FOR ACCOUNTABILITY. YOU  
5 MAY BE PUSHING PEOPLE INTO BEHAVIOR THAT YOU REALLY  
6 DON'T WANT TO BECAUSE IT'S GOING TO UNDERCUT YOUR  
7 INTERPRETATION OF THE SCORES. I'M NOT SAYING THAT  
8 THESE PEOPLE CHEATED OR ANYTHING LIKE THAT. WHAT  
9 I'M SAYING IS WHEN I SEE THESE DATA, I KNOW THAT  
10 THERE'S SOMETHING WRONG. IT'S NOT RIGHT. THERE'S  
11 NOBODY HERE WHO WOULD SAY THAT THERE'S NOT A STRONG  
12 CORRELATION BETWEEN SOCIOECONOMIC STATUS AND TEST  
13 SCORES. WE SEE IT IN EVERYTHING.

14 WHAT HAPPENS WHEN THAT CORRELATION  
15 GOES AWAY AND THEN TWO WEEKS LATER IT POPS BACK UP  
16 AGAIN WITH THOSE SAME STUDENTS? IT'S NOT AS IF  
17 THEY LEARNED SOMETHING AND THEN FORGOT IT. THAT  
18 DOESN'T HAPPEN. AND IT'S NOT IF THE LEVEL CAME UP  
19 OR SOMETHING LIKE THAT. THAT'S WHAT'S UNIQUE ABOUT  
20 THESE DATA. IT'S EXACTLY THE SAME KIDS  
21 ONE-FOR-ONE. HOW COULD THEY SUDDENLY DO SO  
22 POORLY? SO THAT'S WHY I'M SUSPECT ABOUT THE  
23 SCORES. IT'S NOT A CASE OF SOMEBODY COMING AND  
24 SAYING THESE KIDS REALLY DID DO EXCEL, THEY REALLY  
25 DID LEARN A WHOLE LOT, AND SO ON AND SO FORTH, BUT



1 THEY DID COMING IN WITH AN ALTERNATE TEST SHOULD  
2 PRODUCE THOSE RESULTS RIGHT BACK AGAIN? IT DIDN'T  
3 HAPPEN. THAT'S THE PART THAT CONCERNS US.

4 I'M GOING TO -- I KNOW TIME IS SHORT.  
5 I'M GOING TO KEEP MY REMARKS SHORT. DO YOU HAVE A  
6 QUESTION?

7 MS. DAVID: JUST A QUICK ONE BECAUSE I  
8 THINK THE SAME KIND OF PRESSURE IS ON TEACHERS IN  
9 KENTUCKY, AND YET THE SCORES DON'T SHOOT UP.

10 MR. KLEIN: WELL, SCORES HAVE SHOT UP IN  
11 KENTUCKY EVEN THOUGH THE NAEP SCORES DID STAY  
12 PRETTY MUCH --

13 MR. PANKRATZ: THEY HAVE NOT SHOT UP  
14 ANYTHING LIKE IN TEXAS. IN FACT, THEY'VE GONE  
15 BACKWARDS IN MIDDLE SCHOOLS.

16 MS. DAVID: BUT I'M WONDERING IF YOUR  
17 POINTS AREN'T LIMITED TO A PARTICULAR KIND OF  
18 TEST. SOME TESTS ARE EASIER TO PREPARE FOR.

19 MR. KLEIN: WELL, OPEN-ENDED TESTS, QUITE  
20 FRANKLY, IN TERMS OF THE QUESTIONS, ARE MORE  
21 MEMORABLE AND THINGS OF THAT NATURE IF THERE'S A  
22 FORMULISTIC WAY TO SCORE THEM. THE REASON I'M VERY  
23 SENSITIVE TO THIS IS THAT THERE'S A BIG BUSINESS IN  
24 BAR REVIEW COURSES. IT COSTS SOMEBODY ABOUT A  
25 \$1,000 TO TAKE A BAR REVIEW COURSE. AND WHAT ARE



1 THE THINGS THAT THEY TEACH THEM? WELL, THEY'RE NOT  
2 TEACHING THEM CONTENT. THE HAVE CONTENT OUTLINES  
3 AND THINGS LIKE THAT. THEY'RE TEACHING THEM HOW TO  
4 ANSWER THE ESSAY PORTION OF THE EXAM, AND THEN  
5 THERE'S PEOPLE WHO TEACH THEM HOW TO DEAL WITH THE  
6 MULTIPLE-CHOICE PORTION.

7 BUT ON THE ESSAY PORTION, WHEN YOU  
8 PICK UP A BOOK TO READ IT, YOU CAN TELL WHICH BAR  
9 REVIEW COURSE THIS PERSON WHEN TO. IT'S  
10 FORMULISTIC IN THE WAY THEY RESPOND. SO IT'S NOT  
11 AS IF ONE IS EASIER THAN THE OTHER. LET ME TELL  
12 YOU IT'S PROBABLY EASIER TO TEACH THE OPEN-ENDED  
13 STUFF THAN THE MULTIPLE-CHOICE FROM WHAT I'VE  
14 OBSERVED. IT'S PROBABLY EASIER TO DO THAT.

15 BUT WE GET THE SAME RESULTS. THE  
16 RESULTS I SHOWED YOU WERE WITH BOTH MULTIPLE-CHOICE  
17 AND OPEN-ENDED. OKAY.

18 MR. PANKRATZ: IT'S NOT TRUE IN KENTUCKY.  
19 WHEN YOU CAN'T WRITE, YOU CANNOT GET PROFICIENT.  
20 YOU CAN'T EVEN GET APPRENTICE IF YOU CAN'T WRITE.  
21 SO I'M SAYING, IN KENTUCKY AT LEAST, THE BIGGEST  
22 PROBLEM WE HAVE -- WE'D LOVE FOR SOME OF THESE KIDS  
23 JUST TO WRITE IF THEY COULD GET TO THAT FAR. WHAT  
24 I'M SIMPLY SAYING IS YOU REALLY CANNOT FAKE AN  
25 OPEN-RESPONSE ITEM IF YOU HAVE TO WRITE IT AT LEAST



1 FOR THE WRITING PART.

2 MR. KLEIN: WELL, I DON'T THINK YOU CAN  
3 FAKE MULTIPLE CHOICE EITHER IN THAT SAME SENSE.  
4 SEVERAL YEARS AGO, EIGHT OR NINE YEARS AGO, BRIAN  
5 STECHER AND I WERE WORKING ON DEVELOPING A TEST FOR  
6 TEACHERS. AND AS PART OF THAT ACTIVITY, ONE OF THE  
7 THINGS WE WANTED TO DO WAS ASK PROSPECTIVE TEACHERS  
8 TO CONSTRUCT A TEST. HOW GOOD ARE THEY AT  
9 CONSTRUCTING A TEST FOR THEIR CLASSROOM? AND WE  
10 PROPOSED SOME QUESTIONS, AND THE TEACHERS, ALL  
11 CALIFORNIA TEACHERS, LOOKED AT US AND SAID YOU  
12 CAN'T ASK THAT KIND OF QUESTION BECAUSE IT'S NOT  
13 PART OF THE STATE GUIDELINES OF HOW ESSAY QUESTIONS  
14 ARE ASKED. REMEMBER THAT, BRIAN? I FORGET THE  
15 DETAILS OF IT, BUT IT WAS LIKE YOU CAN'T ASK A  
16 PROMPT THIS WAY. A PROMPT IS HIGHLY STYLIZED. AND  
17 YOU GIVE STUDENTS PRACTICE IN ANSWERING PROMPTS  
18 THAT WAY.

19 THIS IS VERY DIFFERENT. I KNOW JIM  
20 POPLIN REAL WELL, I KNOW HOW HE CONSTRUCTS ITEM  
21 SHELLS AND HOW HE GENERATES ITEMS FROM ITEM SHELLS  
22 AND SO ON. AND IF I CAN DO THAT, PRINCETON REVIEW  
23 CAN DO THAT AND OTHER PEOPLE CAN DO THAT TOO,  
24 WHETHER IT'S OPEN-ENDED OR MULTIPLE-CHOICE. I  
25 COULD TEACH SOMEBODY HOW TO DO MY OPEN-ENDED. I'M





1 NOT SUGGESTING THAT I'M IMMUNE TO THAT PROBLEM  
2 EITHER.

3 I'M JUST SAYING WHEN WE THINK ABOUT  
4 TESTING AND CONSTRUCTION OF TESTS, WE SHOULD THINK  
5 ABOUT WHAT IT IS WE WANT TO GENERALIZE TO. WE  
6 REALLY DON'T CARE WHETHER A STUDENT ANSWERS A  
7 PARTICULAR QUESTION RIGHT OR WRONG. WHAT WE CARE  
8 ABOUT IS WHETHER STUDENTS CAN ANSWER QUESTIONS LIKE  
9 THAT. THAT'S WHAT WE WANT TO GENERALIZE TO. WE  
10 WANT TO GENERALIZE TO THIS MUCH LARGER DOMAIN OF  
11 QUESTIONS THAT COULD BE ASKED, OPEN-ENDED,  
12 MULTIPLE-CHOICE. I'D LIKE TO SEE MORE STUFF USING  
13 COMPUTERS WHICH LOOK AT RELATIONSHIPS, SAY, BETWEEN  
14 DECIMALS, FRACTIONS, AND PERCENTS AND UNDERSTANDING  
15 HOW THOSE ARE CONNECTED ON ONE ANOTHER, THINGS LIKE  
16 THAT WHICH WE CAN DO WITH OTHER MODES OF  
17 ASSESSMENT. ALL THOSE ARE OUT THERE, BUT NONE OF  
18 THEM ARE IMMUNE TO THIS PROBLEM.

19 AND SO I THINK WE HAVE TO CHANGE OUR  
20 WHOLE STRATEGY ABOUT HOW WE GO ABOUT DOING TESTING  
21 OF KIDS IN SCHOOLS. I THINK IT'S TIME TO TAKE THE  
22 75-YEAR-OLD METHODOLOGY, WHICH IS ABOUT HOW LONG  
23 WE'VE HAD IT, AND START THINKING ABOUT WHAT WE'RE  
24 GOING TO BE DOING IN THE YEARS AHEAD.

25 MR. FEUER: I JUST WANT TO MAKE SURE I



2 UNDERSTAND THE SEQUENCE HERE IN WHAT YOU PRESENTED,  
3 STEVE, BEFORE I ASK YOU MY QUESTION. YOU GAVE THE  
4 RAND ADMINISTERED TEST, AND THEN TWO WEEKS LATER  
5 THE SAME KIDS TOOK THE TAAS.

6 MR. KLEIN: JUST THE OTHER WAY AROUND.

7 MR. CARNOY: YOU DIDN'T ADMINISTER -- YOU  
8 DIDN'T ADMINISTER THE TAAS, RIGHT?

9 MR. KLEIN: NO. WE DIDN'T ADMINISTER THE  
10 TAAS. WE ADMINISTERED THE RAND TEST.

11 MR. FEUER: ON THE RAND TEST YOU FIND THE  
12 MORE PREDICTABLE, MORE EXPECTED KIND OF RESULT.

13 MR. KLEIN: TYPICAL.

14 MR. FEUER: AND ON THE TAAS YOU FIND --

15 MR. KLEIN: DOESN'T MAKE A DIFFERENCE. WE  
16 ALSO DO STANFORD IX. WAS IT STANFORD IX THAT WE  
17 USED AT THIS SITE FOR THE MULTIPLE-CHOICE, BRIAN?  
18 I'M TRYING TO REMEMBER. IT WAS ONE OF THE  
19 STANDARD -- IT WAS EITHER THE ITBS OR THE STANFORD  
20 IX MULTIPLE-CHOICE SCIENCE TEST.

21 MR. STECHER: YOU SHOULD SAY THAT THE  
22 MULTIPLE-CHOICE ONE WAS NOT A RAND DEVELOPED TEST,  
23 BUT THAT WAS A STANDARD OFF-THE-SHELF TEST OF  
24 SCIENCE.

25 MR. KLEIN: RIGHT.

MR. FEUER: SO YOU RULE OUT THE

1 POSSIBILITY THAT THE TESTS ARE IN SOME FUNDAMENTAL  
2 WAY PICKING UP ON A COMPLETELY DIFFERENT SORT OF  
3 COGNITIVE LOADINGS AND TASKS AND THINGS.

4 MR. KLEIN: TESTS ARE DIFFERENT, MIKE.  
5 I'M NOT SAYING THE STANFORD IX IS THE SAME AS THE  
6 ITBS; BUT IF YOU LOOK AT READING AND WHAT PEOPLE'S  
7 CONCEPTION ARE, HOW DIFFERENT ARE THEY?

8 MR. FEUER: WELL, WHEN --

9 MR. KLEIN: ARE THEY DIFFERENT ENOUGH TO  
10 GET A CORRELATION FROM .7 DOWN TO .0?

11 MR. FEUER: I'M NOT SAYING THAT. BUT THE  
12 QUESTION IS IS READING READING? THIS IS A GOOD  
13 PLACE TO BE ASKING THAT QUESTION BECAUSE CALIFORNIA  
14 HAS, I THINK, HAD TO ADDRESS THAT. BUT YOU'RE  
15 RULING OUT THE POSSIBILITY THAT THE TESTS ARE  
16 DIFFERENT ENOUGH TO EXPLAIN THE DIFFERENCE IN  
17 RESULTS. I'LL GO ALONG WITH YOU ON THAT.

18 THE ONLY THING THAT YOU HAVEN'T  
19 BROUGHT UP AS A POSSIBILITY, THAT SOMEHOW A SIGNAL  
20 WAS CONVEYED, A MESSAGE WAS CONVEYED TO THE KIDS OF  
21 A FUNDAMENTALLY DIFFERENT NATURE IN THESE TWO  
22 ADMINISTRATIONS. WHAT I'M THINKING HERE, AND YOU  
23 CAN TELL ME IF YOU THINK THIS IS OFF THE WALL, BUT  
24 YOU KNOW, CLAUDE STEELE HAS THIS EVIDENCE ABOUT  
25 WHAT HAPPENS TO THE RESULTS WHEN KIDS GET THE SENSE



1 THAT THE RESULTS ARE SOMEHOW GOING TO BE USED TO  
2 EITHER CONFIRM OR REFUTE A STEREOTYPE OR BIAS ABOUT  
3 THEIR RACIAL OR ETHNIC GROUP.

4 I CAN IMAGINE TEACHERS ADMINISTERING  
5 THE TAAS AND CONVEYING A SOMEWHAT MORE POSITIVE  
6 MESSAGE ABOUT THE OVERALL EXPERIENCE THAN MIGHT BE  
7 THE CASE IN A MORE STANDARD ADMINISTRATION. DO YOU  
8 THINK THAT HAS ANY CREDENCE AT ALL?

9 MR. KLEIN: I REALLY DON'T THINK SO IN  
0 THIS SITUATION FOR A COUPLE OF REASONS. ONE IS  
1 THAT WE DID FIVE DAYS OF TESTING THESE KIDS. FIVE  
2 CLASSROOM PERIODS PER KID OF TESTING. THE KIDS  
3 LOVED IT BECAUSE IT WAS THIS HANDS-ON STUFF. THEY  
4 CAN PLAY WITH PENDULUMS AND LEVERS AND CARS AND  
5 TRUCKS AND ANIMALS AND ALL SORTS OF STUFF BECAUSE  
6 FOR FOUR OF THE FIVE CLASSROOM PERIODS THERE WAS  
7 ALL THIS OPEN-ENDED STUFF, AND THEY WERE VERY HAPPY  
8 ABOUT THIS.

9 I DON'T THINK -- THE SITUATION WAS  
0 REALLY DIFFERENT. THE OTHER THING IS THAT IF THE  
1 KIDS WERE MARKING RANDOMLY OR SOMETHING ELSE WAS  
2 GOING ON, THE RELIABILITY WOULD GO DOWN THE TUBES.  
3 WE DIDN'T SEE THAT AT ALL. RELIABILITIES WERE  
4 FINE. AND IT WAS NO DIFFERENT ON OUR TEXAS SITES  
5 THAN IT WAS IN OUR OTHER PLACES.



1 AND LIKE I SAID, IT WASN'T SOMETHING  
2 THAT WAS UNIQUE TO TEXAS. I'M NOT SUGGESTING THAT  
3 AT ALL, BUT IT WAS HERE AND NOT THERE AND THAT KIND  
4 OF THING, AND IT WASN'T AT EVERY SCHOOL BECAUSE IT  
5 LOOKED TO ME LIKE WE GOT TWO CURVES GOING ON  
6 THERE.

7 MR. POGGIO: PUT UP THE LAST ONE AGAIN.  
8 DID YOU GO IN AND TALK TO THE SCHOOLS AT THAT POINT?

9 MR. KLEIN: WE DIDN'T HAVE THE DATA AT  
10 THAT POINT. THE OLD CLASSIC, TAKES YOU NINE MONTHS  
11 TO GET THE DATA ANALYZED. SEE IF I CAN FIND IT.  
12 HERE'S THE ONE WITH THE CIRCLE -- CLUSTER, THERE'S  
13 ACTUALLY EIGHT DOTS, BUT TEN SCHOOLS RIGHT UP IN  
14 HERE.

15 MR. POGGIO: I'M WONDERING TO WHAT EXTENT  
16 THE PROBLEM IS WITH THE ONES THAT WOULD CONFIRM THE  
17 THINGS YOU EXPECTED. WELL, THOSE TWO IN THE LOWER  
18 RIGHT-HAND CORNER.

19 MR. KLEIN: THESE HERE?

20 MR. POGGIO: GO TO THE LOWER RIGHT-HAND  
21 CORNER AND PICK UP THE TWO THAT ARE THERE, THAT ONE  
22 AND THE ONE JUST ABOVE IT. WHAT I'M SAYING IS YOU  
23 OUGHT TO BE TALKING TO THOSE TWO AND SEE IF THEY'RE  
24 DOING ANYTHING AND IF WHAT YOU'RE GETTING THERE IS  
25 WE'RE REALLY NOT PAYING ATTENTION TO TAAS. I DON'T



SAY YOU HAVE TO GO LOOK AT THE OTHER EIGHT.

MR. KLEIN: THIS IS NOT TAAS. THIS IS TAAS. THIS IS THE TAAS STATE MULTIPLE-CHOICE TEST.

MR. POGGIO: WHAT I HEAR YOU SAYING IS YOU'RE WILLING TO TAKE AS FACT THE TWO THAT APPEAR TO BE OUTLIERS TO QUANTIFY YOUR DISTRIBUTION.

MR. KLEIN: THE TWO THAT ARE OUTLIERS, THIS GROUP OF TEN?

MR. TRENT: THE BOTTOM TWO.

MR. FEUER: YOU'VE STILL GOT A U-SHAPED CURVE WE CAN'T EXPLAIN.

MR. KLEIN: YOU'VE GOT A U-SHAPED CURVE NO MATTER HOW YOU LOOK AT IT.

MR. POGGIO: THE QUESTION IS IS SOMETHING IRREGULAR TAKING PLACE IN THE UPPER RIGHT-HAND CORNER? AND IT SEEMS TO ME WE CAN NOTE SOMETHING, BUT THE ANSWER IS THROUGH INTERVIEWS AND OBSERVATIONS AT THIS POINT RATHER THAN A DEDUCTION THAT SAYS THERE'S A PRACTICE THAT'S GOING ON THAT'S DECEITFUL.

MR. KLEIN: WHAT I SAID WAS THERE'S MANY POSSIBLE EXPLANATIONS. YOU'RE LISTENING. WHAT I SAID WAS THERE ARE MANY POSSIBLE EXPLANATIONS TO WHAT IS GOING ON HERE. RIGHT? DID I SAY THAT?

MR. POGGIO: RIGHT.



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25

MR. KLEIN: DID I USE THE WORD "DECEIT"?

MR. POGGIO: NO. WHAT YOU DESCRIBED WAS TEACHERS GOING UP AND DOWN THE AISLES.

MR. KLEIN: I SAID A BREAK IN STANDARDIZATION.

MR. POGGIO: I DIDN'T HEAR YOU TALK ABOUT THE OTHER TWO AS SCHOOLS WHERE THERE WAS A LACK OF ATTENTION TO THE GOALS IN THE PROGRAM. YEAH, I'M MAKING AN INFERENCE.

MR. COLE: WHAT GRADE WERE THESE TAAS?

MR. KLEIN: THIS IS GRADE FIVE, I BELIEVE. BRIAN, THIS IS GRADE FIVE, RIGHT?

MR. STECHER: YES.

MR. COLE: YOURS IS SCIENCE, WHEREAS THESE ARE READING?

MR. KLEIN: YEAH. WE HAD MATH AND SCIENCE AT THIS SITE. WE HAD MATH AND SCIENCE BOTH. THAT'S WHY WE HAD ALL THOSE MEASURES UP THERE.

MR. JOHNSON: SO DID YOU GIVE A READING ASSESSMENT OTHER THAN TAAS?

MR. KLEIN: NO.

MR. JOHNSON: YOU ARE SAYING THAT BECAUSE YOU HAD A PREDICTABLE CURVE ON A MATH ASSESSMENT AND A PREDICTABLE CURVE ON A SCIENCE ASSESSMENT, YOU FIND IT SURPRISING TO FIND THAT STUDENTS WOULD



PERFORM AS THEY PERFORMED ON A READING ASSESSMENT?

MR. KLEIN: WE ALSO GOT THE SAME THING IN TAAS MATH.

MR. JOHNSON: THEN ON THAT MATH ASSESSMENT, DID YOU --

MR. KLEIN: HERE'S TAAS MATH.

MR. JOHNSON: AND SO FOR THE MATH ASSESSMENT, IN COMPARING THE RAND ASSESSMENT AND THE TAAS ASSESSMENT, DID YOU LOOK AT THE CONTENT THAT WAS BEING ASSESSED, THE OBJECTIVES THAT WERE BEING ASSESSED TO DETERMINE THE CONSISTENCY BETWEEN THE TWO INSTRUMENTS?

MR. KLEIN: WE DID NOT; HOWEVER, I CAN REFER YOU TO DICK YEAGER WHO LOOKED AT THAT SAME SORT OF PROBLEM IN OTHER AREAS TO SEE WHAT THE CONNECTION WAS BETWEEN THE OBJECTIVES AND THE CONTENT AND FOUND THAT THAT DOESN'T EXPLAIN IT. IT MIGHT IN THIS SITUATION. I DON'T THINK SO.

MR. JOHNSON: FOR YOU TO MAKE THAT ASSERTION WITHOUT HAVING THAT KIND OF EVIDENCE DOESN'T HELP US BECAUSE YOU ARE MAKING -- THE ASSERTION THAT YOU ARE MAKING IS EXTREMELY IMPORTANT. YOU ARE SAYING THAT THOSE SCHOOLS THAT ARE DOING WELL, THAT MAYBE HAVE SPENT CONSIDERABLE TIME AND ENERGY ON IMPROVEMENT, YOU'RE, IN ESSENCE,





1 DENIGRATING THOSE EFFORTS BASED ON A HALF ANALYSIS  
2 OF THE FACTS.

3 MR. KLEIN: WELL, YOU ARE ENTITLED TO YOUR  
4 OPINION. OKAY. YOU ARE ENTITLED TO YOUR OPINION.  
5 I THINK THE DATA, WHEN YOU LOOK AT A .04  
6 CORRELATION BETWEEN SOMETHING THAT EVERYBODY IS  
7 COMPLAINING ABOUT FOR YEARS AND YEARS, WHICH IS THE  
8 CLOSE CORRELATION BETWEEN SOCIOECONOMIC STATUS AND  
9 MULTIPLE-CHOICE TEST SCORES, PEOPLE -- IN YOUR BOOK  
10 THERE'S TONS OF ARTICLES. I AM LOOKING THROUGH IT  
11 THIS MORNING. HERE'S TONS OF ARTICLES, NEWSPAPER  
12 ARTICLES AND SO ON, WHICH TALK ABOUT THE FACT THAT  
13 TEST SCORES ARE HIGHLY RELATED TO SOCIOECONOMIC  
14 STATUS. DOES ANYBODY DISAGREE?

15 MR. JOHNSON: IF THAT SCATTER PLOT IS THE  
16 PROBLEM, THEN I THINK THAT -- WELL, I'LL JUST  
17 SUGGEST THAT THAT SCATTER PLOT IS NOT THE PROBLEM.  
18 THE PROBLEM IS THE SCATTER PLOT THAT IS THE  
19 PREDICTABLE LINE. AND WHAT THE TEXAS REFORM IS  
20 ABOUT IS ABOUT CHANGING THAT PREDICTABILITY SO THAT  
21 STUDENTS IN HIGH POVERTY SCHOOLS, STUDENTS IN HIGH  
22 MINORITY SCHOOLS WILL ACHIEVE AT HIGHER LEVELS.  
23 THAT'S THE EXPRESSED INTENT OF THE SYSTEM, AND  
24 THERE'S EVIDENCE THAT THE SYSTEM HAS ACHIEVED THAT  
25 RESULT. AND YOU ARE NOT GIVING US EVIDENCE THAT



THE SYSTEM HAS NOT ACHIEVED THAT RESULT.

MR. KLEIN: I DIDN'T SPEAK TO A SYSTEM THAT ACHIEVED THAT RESULT OR NOT. I'M JUST SAYING I DON'T TRUST THE SCORES BECAUSE I SEE RELATIONSHIPS THAT ARE COUNTERINTUITIVE AND COUNTEREXPERIENTIAL.

MR. JOHNSON: WASN'T YOUR FUNDAMENTAL OBJECTIVE BEGINNING THIS PROCESS, DOING THIS COMPARISON, WHAT DID YOU GO IN TO DO, TO PROVE, NOT PROVE, TO SEE?

MR. KLEIN: IT'S TOTALLY UNRELATED TO THIS. WHAT WE WERE INTERESTED IN WAS TAKING A LOOK AT WHETHER THE EFFECTS OF THE REFORM PRACTICES IN SCHOOLS WOULD BE MORE LIKELY TO BE SHOWN ON OPEN-ENDED, HANDS-ON KINDS OF TESTS THAN THEY WOULD BE ON MULTIPLE-CHOICE TESTS. THOSE OF YOU WHO KNOW ME WELL KNOW THAT I'M A STRONG PROPONENT OF OPEN-ENDED MEASURES WHEN USED IN CONJUNCTION WITH OTHER KINDS OF THINGS. AND I THOUGHT, WELL, GEE. WHAT'S GOING TO BE MORE SENSITIVE TO THE EFFECTS OF THE REFORM? IS IT GOING TO BE A STANDARD MULTIPLE-CHOICE TEST, OR IS IT GOING TO BE SOMETHING THAT'S MORE OPEN-ENDED AND ALLOWS STUDENTS TO DEMONSTRATE THEIR UNDERSTANDING AND SO ON AND SO FORTH? THAT WAS THE MOTIVATION. SO WE



1 WANTED BOTH KINDS OF MEASURES IN THIS STUDY.  
2 THAT'S WHY WE GAVE THEM BECAUSE, LO AND BEHOLD,  
3 TEXAS WASN'T.

4 SO PART OF THE PROJECT WAS WE WOULD  
5 GIVE MEASURES THAT WOULD SUPPLEMENT WHAT TEXAS  
6 WASN'T DOING. AND THE REASON WE GAVE A SCIENCE  
7 TEST WAS BECAUSE THEY WEREN'T GIVING A SCIENCE TEST  
8 AT THAT PARTICULAR GRADE LEVEL OR WHATEVER THE CASE  
9 MIGHT BE. IT DEPENDED ON THE SITE. THAT WAS OUR  
10 MOTIVATION, UNRELATED TO THIS ISSUE. IT WAS JUST A  
11 FIND.

12 THAT'S WHY WE WEREN'T LOOKING AT THE  
13 QUESTION OF WHAT IS THE RELATIONSHIP BETWEEN THE  
14 OBJECTIVES MEASURED BY THIS AND THAT TEST AND SO  
15 ON. BUT ANYBODY WHO HAS WORKED IN THE FIELD KNOWS  
16 THAT IF I GO IN AND I GIVE THE CTBS OR THE STANFORD  
17 IX OR THE ITBS OR WHATEVER ALPHABET TEST YOU WANT  
18 TO GIVE, THE CORRELATIONS ARE GOING TO BE PRETTY  
19 HIGH AMONG THOSE MEASURES UNLESS THERE'S SOMETHING  
20 UNUSUAL GOING ON.

21 THAT'S WHAT WE FIND. YOU CAN LOOK --  
22 DICK YEAGER DID THAT STUDY. YOU CAN TALK TO A  
23 NUMBER OF PEOPLE WHO HAVE DONE THOSE KINDS OF  
24 STUDIES. FOLKS, WHILE READING MAY NOT BE READING,  
25 IT'S STILL A LOT MORE ALIKE ONE ANOTHER.



2 MR. CARNOY: STEVEN, CAN I ADJUDICATE  
3 SOMETHING HERE? THE GENERAL IDEA OF ALL THE STATE  
4 REFORMS, THE STRONG EVALUATIONS, I THOUGHT, WAS TO  
5 TRY TO LAY IT TO RAISE EVERYBODY'S LEVEL OF  
6 PERFORMANCE. IS THAT INCORRECT? RAISE THE CURVE.  
7 OKAY. AND HOPEFULLY TO TILT THE CURVE SOMEWHAT.

8 I DON'T THINK ANYBODY EVER EXPECTED,  
9 GIVEN, YOU KNOW -- I THOUGHT WHAT WAS TRYING TO BE  
10 SOLVED HERE WAS THE FACT THAT EVEN THOUGH KIDS COME  
11 IN SCHOOL WITH DIFFERENCES ALREADY, QUITE HIGH, IN  
12 FACT, ESPECIALLY WITH THESE TESTS, THAT APPARENTLY  
13 THESE DIFFERENCES SPREAD, PARTICULARLY IN -- UP TO  
14 ABOUT HIGH SCHOOL. THAT'S MY UNDERSTANDING OF SOME  
15 OF THE RECENT DATA.

16 SO THE IDEA WAS AT LEAST TO CLOSE THE  
17 GAP ON THE PART THAT THE SCHOOL WAS NOT DOING WELL.  
18 THAT WOULD ALREADY BE A TREMENDOUS ACCOMPLISHMENT.  
19 NOW, SO IT'S HARD TO DO THAT. IN OTHER WORDS, TO  
20 TILT -- TO LOWER THE CORRELATION SOMEWHAT. AND  
21 WHAT YOU'RE ARGUING IS THAT THE TEST IN THESE  
22 SCHOOLS BASICALLY ELIMINATED THE CORRELATION  
23 TOTALLY.

24 MR. KLEIN: TURNED LEAD INTO GOLD.

25 MR. CARNOY: BUT IF YOU DESIGNED AN EASY  
ENOUGH TEST, COULDN'T YOU DO THIS AFTER A WHILE?



1 I'M JUST ASKING.

2 MR. KLEIN: MAKING THE TEST EASIER WON'T  
3 DO IT.

4 MR. HAKUTA: BUT IF YOU HAVE A CEILING  
5 EFFECT.

6 MR. KLEIN: IF EVERYBODY HAD THE SAME  
7 SCORE, YEAH. IF YOU TOOK IT TO THE EXTREME, YES.  
8 THAT'S NOT HAPPENING HERE.

9 MR. POGGIO: IT'S NOT HAPPENING IN FREE  
10 AND REDUCED LUNCH. WE OBVIOUSLY HAVE A RESTRICTION  
11 OF RANGE THAT'S NEEDED -- YOU CAN SOMEHOW  
12 DISINTENUATE THAT. I DON'T KNOW HOW YOU GO BEYOND  
13 A HUNDRED PERCENT.

14 MR. CARNOY: THE CHANCES OF GETTING ONE  
15 WRONG ON THIS TEST, JUST BY MAKING A MISTAKE ON A  
16 MULTIPLE-CHOICE TEST, NO MATTER HOW SMART YOU ARE,  
17 NO MATTER HOW WELL YOU KNOW THE SUBJECT, YOU KNOW  
18 BETTER THAN I, BUT I THINK THEY'RE REASONABLY  
19 HIGH. SO TO GET A 90 PERCENT, I DON'T KNOW HOW  
20 MANY QUESTIONS ARE ON THE TAAS MATH TEST.

21 MS. SMISKO: DEPENDS ON THE GRADE LEVEL,  
22 BUT AT GRADE FIVE IT'S PROBABLY IN THE NEIGHBORHOOD  
23 OF 60.

24 MR. KLEIN: SIXTY ITEMS WHICH IS A PRETTY  
25 GOOD RANGE. IN TERMS OF RESTRICTION RANGE, YOU GO



1 FROM ABOUT 20 PERCENT OF THE KIDS UP TO A HUNDRED  
2 PERCENT. IT IS SOME RESTRICTION, NOT MUCH. IT'S  
3 CERTAINLY NOT CURTAILED, CENSORED DATA OR SOMETHING  
4 LIKE THAT. WE'VE CUT IT OFF AT A CERTAIN POINT OR  
5 SOMETHING LIKE THAT.

6 MR. CARNOY: WHAT'S INTERESTING IS THAT  
7 THE SCORES ARE PRETTY LOW, MUCH LOWER THAN THE  
8 AVERAGE IN THE STATE.

9 MR. FEUER: I THINK THE MESSAGE HERE IS A  
0 SLIGHTLY DIFFERENT ONE, IF I MIGHT. AND THAT IS  
1 THAT I THINK STEVE WOULD AGREE THAT THE OBJECTIVE  
2 HERE IS THE WAY YOU'VE DESCRIBED IT, MARTIN; THAT  
3 IS, TO AT LEAST TRY TO CORRECT THIS AWFUL DOWNWARD  
4 SLOPING FUNCTION. AND IT IS SUCH AN IMPORTANT  
5 OBJECTIVE, THAT IF THERE'S EVIDENCE THAT IT'S  
6 HAPPENING, WE WANT TO BE SURE THAT THE EVIDENCE IS  
7 CREDIBLE.

8 NOW, WHEN YOU LOOK AT NAEP SCORES  
9 OVER A 20- OR 25-YEAR PERIOD AND YOU OBSERVE THAT,  
0 IN FACT, THE GAP IS SHRINKING AND THAT MINORITY  
1 GROUPS ARE DOING BETTER, YOU COME AWAY FROM THAT  
2 SOMEWHAT HEARTENED THAT THERE HAS BEEN PROGRESS.  
3 AND YOU ARE HEARTENED BECAUSE, LIKE IT OR NOT, NAEP  
4 HAS CREDIBILITY FOR DOING THAT KIND OF THING.

5 THE HOPE, THEN, IS THAT THESE RESULTS



1 HAVE THAT KIND OF CREDIBILITY. AND WHAT STEVE IS  
2 SUGGESTING HERE IS THAT WE BETTER GET REAL CAREFUL  
3 AND METICULOUS ABOUT, YOU KNOW, THE CREDIBILITY OF  
4 THESE KINDS OF RESULTS.

5 THE QUESTION OF WHAT MIGHT HAVE GONE  
6 WRONG IN THE CLASSROOMS OR WHETHER THE SCHOOLS DID  
7 OR DID NOT EMPHASIZE TO THE KIDS CERTAIN THINGS,  
8 THESE ARE, YOU KNOW, THESE ARE CONJECTURES, BUT I  
9 DON'T THINK -- FOR ME THE FUNDAMENTAL POINT IS THAT  
10 ONE WANTS TO TAKE THESE DATA SERIOUSLY. AND WHEN  
11 YOU SEE A CORRELATION DROP TO .04, IT DOES GIVE  
12 PAUSE. BUT THE FACT THAT IT IS COUNTERINTUITIVE,  
13 YOU SEE THIS IS WHERE I SORT OF PART WITH YOU  
14 BECAUSE IT IS PRECISELY THE OBJECTIVE OF THIS IS TO  
15 BREAK THOSE INTUITIONS.

16 MR. KLEIN: I DON'T DISAGREE WITH THE  
17 GOAL. WE WANT TO MOVE EVERYBODY UP. WE'D LIKE TO  
18 CLOSE THE GAP. THERE'S NO DISAGREEMENT. IT'S WHEN  
19 SUDDENLY IT HAPPENS AND THEN TWO WEEKS LATER IT  
20 DISAPPEARS, THAT'S THE PROBLEM. IN OTHER WORDS, I  
21 WOULD BE MUCH MORE CONVINCED THAT YOU ACCOMPLISHED  
22 YOUR GOALS IF I CAME IN AND GAVE A TEST THAT WAS  
23 SIMILAR IN NATURE AND I GOT THE SAME RESULT THAT  
24 YOU DID. WE HAD TALKED ABOUT THAT RIGHT BEFORE THE  
25 BREAK, RIGHT.



1                   THE PIECE THAT'S DIFFERENT HERE IS I  
2 TESTED THE SAME KIDS TWO WEEKS LATER AND IT  
3 DISAPPEARED. WHERE DID IT GO? IT WENT INTO THIN  
4 AIR. HE SAYS, WELL, MAYBE IT'S BECAUSE THE  
5 OBJECTIVES ARE DIFFERENT. WELL, IT'S A  
6 POSSIBILITY, BUT I DOUBT THAT IT COULD EXPLAIN ALL  
7 OF THIS OR MOST OF IT. IT'S EXTREMELY UNLIKELY. I  
8 THINK THAT YOU HAVE TO GIVE ME THAT, THAT IT'S  
9 EXTREMELY UNLIKELY. IF YOU WANT TO BET, I'LL BET  
10 YOU.

MR. FEUER: THAT WOULD REQUIRE THAT YOU DO  
SOMETHING LIKE AN EQUATING STUDY OF THESE TESTS.

MR. KLEIN: WHAT IT WOULD SUGGEST IS THAT  
MAYBE TEXAS SHOULD DO ITS OWN AUDITING TEST.

MR. COLE: WAS YOUR TEST TIED IN ANY WAY  
TO THE <sup>TEKS</sup> PEAKS? DID IT HAVE A CORRELATION WITH THE  
TEXAS --

MR. KLEIN: NO. WE USED OFF-THE-SHELF  
TESTS. THESE ARE STANDARD, OFF-THE-SHELF TESTS  
WITH THE EXCEPTION OF OUR HANDS-ON MEASURES, WHICH  
CONFORM WITH -- I WOULD GUARANTEE ANY REFORM SET OF  
STANDARDS YOU HAVE, THOSE OPEN-ENDED MEASURES WOULD  
FIT. I'M SURE THEY WOULD FIT.

MR. COLE: UNLESS THEY WERE TEACHING A  
DIFFERENT AREA OF CURRICULUM AT THAT GRADE LEVEL

