dallasnews.com
The Dallas Morning News

Common questions about analyzing tests for cheating

02:35 AM CDT on Sunday, June 3, 2007

Can you really detect cheating through statistics?

The science of detecting too-similar answer sheets is decades old and well accepted among psychometricians – the people who design standardized tests. It has been the subject of dozens of academic papers in respected journals. Statistical methods similar to *The News*' are used to detect cheating on major national tests like the SAT and are sometimes used to invalidate a test taker's scores. At least one university (McGill in Montreal) has invalidated scores on final exams based on a statistical cheating analysis. Different statistical methods generally flag the same students for their suspicious scores. And experts say the methods are reliable windows into the scale of cheating on school campuses.

One weakness of the statistical analysis is that it cannot detect in what direction the cheating occurred. In other words, if Johnny and Jimmy are flagged, the analysis cannot tell which one cheated off of the other. Ideally, in an investigation, schools would have supporting evidence, such as a seating chart showing that a flagged pair sat next to one another on test day. Current Texas testing rules do not require schools to keep seating charts – or even to record which classrooms students took the TAKS in.

How do we know to trust the data? Couldn't you just be flagging random kids? Isn't there a risk of a false positive?

That's always a risk. Looking for cheaters is a bit like DNA analysis: It can't identify a match with 100-percent certainty. It can only say that the chance of a false positive is very, very small. *The News* analysis was designed to minimize that chance. It was based on the detection methodology of cheating researcher George Wesolowsky and used very conservative assumptions. According to Dr. Wesolowsky, those assumptions should result in a completely innocent school being falsely flagged less than once out of every 10,000 cases.

Here's another test of the effectiveness of the methodology. At one point, Dr. Wesolowsky purposefully entered answer sheets from more than 100 different schools into his computer program – without telling it which students went to which schools. He then asked his program to determine which pairs of students had answer sheets that suggested cheating

If the program was flagging kids willy-nilly – that is, if it wasn't catching cases of true collusion between students or adults – you'd expect only a small fraction of the pairs it found to be from within the same school. But that wasn't the case. The program flagged 8,548 different pairs of students out of that data. Of those, only 57 featured students from different schools. In other words, without knowing where students were, the program flagged pairs within the same school 99.3 percent of the time.

(None of those cross-school pairs are included in the *News* analysis – although many of them connected pairs of students from nearby schools, leaving open the possibility that text messaging was used to cheat in those cases.)

Couldn't these kids have all the same answers because they studied together? Or couldn't they have had a bad teacher who taught them all the wrong answers?

Experts say those aren't valid reasons for the sort of identical answers found in the *News* study. First, kids study together in every Texas school – but two-thirds of all Texas schools had not even a single student flagged for cheating. If studying together led to flagging, you'd expect flagging to be much more common than it is. In fact, a number of studies have found that studying together does not actually lead to markedly increased similarity among students' answer sheets.

Third, if teachers were teaching the material incorrectly, you'd expect the entire class (or close to it) to get those questions wrong. That's not true in the vast majority of cases found in the *News* analysis. The most common form of cheating entailed a small group of students who had identical wrong answers that differed significantly from the rest of their class.

Perhaps most importantly, in cheating analyses of places where seating charts were kept, statistical detection correlates strongly with where students sit. In other words, studying together or improper teaching don't lead to flags – but sitting together does.

How much of this is the teachers' fault?

It's impossible to know for sure. Most of the cheating detected in the *News* study linked isolated individual students to one another – the sort of results you'd expect from one kid sneaking answers off a buddy or a neighbor. On the other extreme are a few schools – mostly charter schools – where a majority of students' answer sheets are identical or nearly identical. That's a pattern that could easily be the result of adult involvement.

In between, you have a substantial number of schools – over 100 – where cheating was not systemic, but wasn't isolated to a few pairs of students either. These are schools where 10 percent or more of a school's answer sheets were flagged. In those schools, there would appear to be serious problems with test security. That could be the result of sloppy work by adults, or it could be intentional.

Is Texas an extreme case? Do we cheat more than other places?

It's difficult to know, since this statewide analysis is one of the first of its kind. And Caveon – the company which has performed this sort of analysis in Texas and North Carolina – doesn't publicize its findings. But in most schools, according to researchers, the cheating levels fall within an expected range. In many surveys, a majority of teens admit cheating at least occasionally on school tests.

Some of the most extreme cases, however, are – well, extreme. Several researchers said they had never seen the scale of cheating witnessed in the most flagrant Texas schools, like Houston's Jesse Jackson Academy.

Didn't the Texas Education Agency clear many of these schools, saying there was no cheating?

Yes. But the agency's investigation, in a majority of cases, consisted only of asking school officials to complete a questionnaire detailing their test-security procedures. In a smaller number of cases, TEA sent investigators to schools to interview staff members about the 2005 test, which by that point had occurred more than a year and a half earlier. They did not use any student-level testing data. Thus far, all of the schools whose investigation TEA has completed have been cleared.