

Cautions about Rating BC's Schools

Philip H. Winne, John C. Nesbit, Carmen L. Z. Gress

Faculty of Education, Simon Fraser University

The Issue

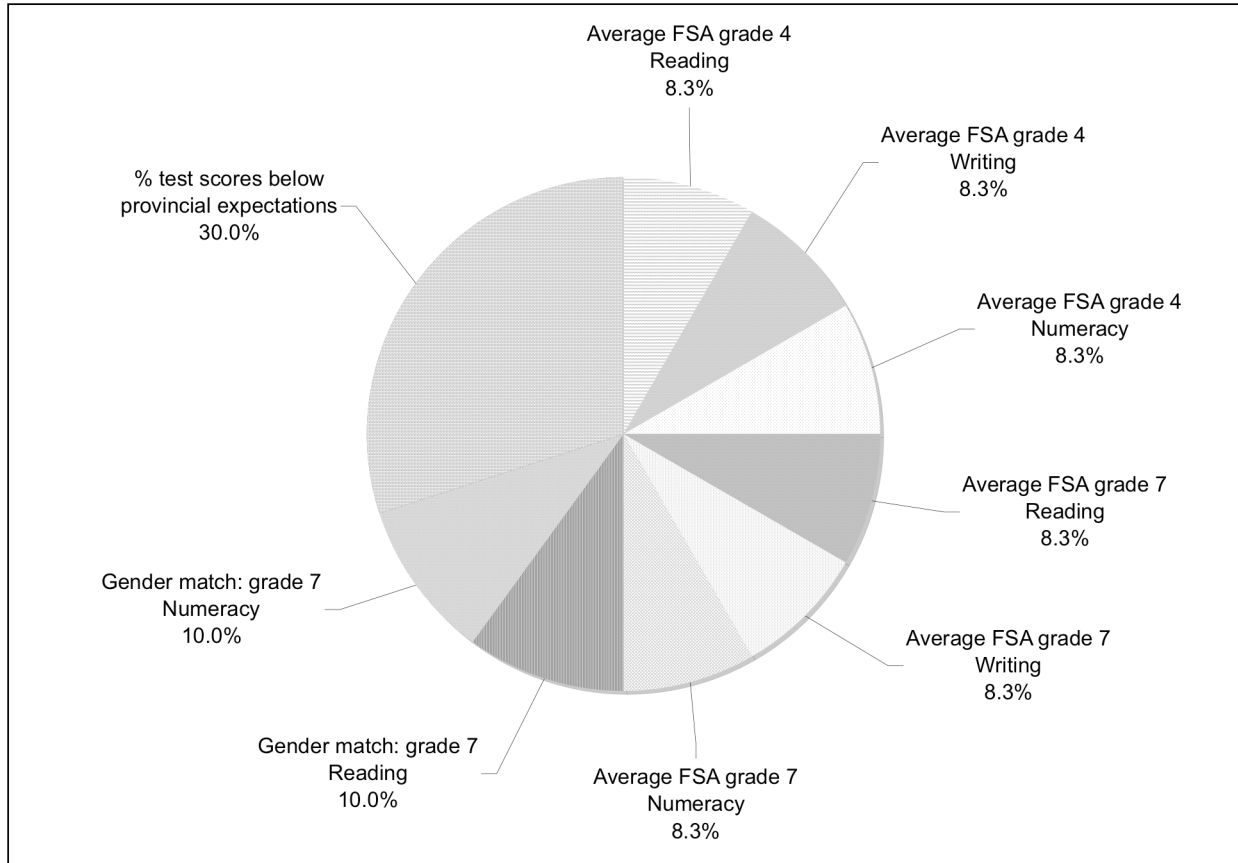
This is the second year *The Vancouver Sun* has published a special section on the academic effectiveness of BC's elementary schools as rated by the Fraser Institute. We're told the Institute's ratings of elementary schools, as well as the report it released in April rating secondary schools, are widely discussed. Reportedly, families consult them when buying homes in hope of boosting educational opportunities for their children. Although it doesn't happen in BC, in the U.S., some jurisdictions use ratings like these, along with other information, to decide how much funding schools receive.

It's worth keeping in mind the Institute's rating of a school is not the same thing as what students know or how competent teachers are or how effective schools are. Focusing on students, there's more to what they know than any one rating can reveal. As well, there is evidence that ratings like these are related to socioeconomic status and wealth. For example, see Selcuk Sirin's award winning article, *Socioeconomic status and academic achievement: A Meta-analytic review of research 1990-2000*, published in the *Review of Educational Research* in 2005, and the 2006 Statistics Canada study, *Income and the Outcomes of Children*, by Shelley Phipps and Lynn Lethbridge, respectively. When important decisions are at stake, it's important to understand what these kinds of ratings are and what limits they have.

The Institute poses a very worthwhile question: "In general, how is the school doing, academically?" To answer it, they calculate a rating from 0 to 10 points for each elementary and secondary school that enrolls at least 15 students. Our answer to this question would take a book. We give just a few highlights here, using data on BC's elementary and secondary schools we obtained through a research agreement with the Ministry of Education to explore issues affecting these kinds of ratings.

What Are the School Ratings?

Ratings, like the Institute's, indicate or point at something. The Institute intends to point at how well a school is doing academically, in general. In technical terms, the Institute's 10-point rating is a composite indicator. For BC's elementary schools, the Institute's rating puts together nine ingredients as shown in the pie chart. All ingredients are based directly on scores students earn on Foundation Skills Assessment (FSA) exams, an annual province-wide assessment of reading, writing, and numeracy skills in grades four and seven.



Deciding what parts to include in a composite indicator calls on judgment and values. For example, if males and females score differently on FSA exams—a gender gap instead of a gender match—it’s worth investigating why. If a gender gap was caused by something schools could control, some would say that the schools should strive to close the gap. Simply closing a gender gap might not be a good thing. If the gender gap became a perfect gender match by decreasing scores of the group that initially scored higher while scores of the group that initially scored lower were not increased, that would be a step backwards. Others might prefer to focus on increasing average performance rather than closing a gap.

We won’t debate very much about the parts the Institute uses in its composite indicator. We could debate whether the percents attached to each part are appropriate. But we won’t do that either. We choose to explore other issues.

Problems with Ratings of Every Kind

There always is some error in measurements like those used to create the Institute’s rating. We can tolerate some error, so the question is whether the level of accuracy is high enough for particular purposes. The tire pressure gauge at the gas station is accurate enough for driving but it’s probably not precise enough for an engineer.

The technical term for error in a measurement is unreliability. Because every part of the Institute’s composite rating of a school has some unreliability, the rating does, too. The question

is whether the composite ratings are precise enough to use for practical matters like choosing a school.

Suppose an elementary school has an Institute rating of, say, 6.2 – that’s the provincial average using the Fraser Institute’s sample of schools. How much higher or lower might this rating actually be? In theory, it should be possible to calculate this directly and quite accurately. But, because some critical information is not collected, no one can do this. We can, however, come close.

For the sake of debate, we describe the unreliability of the Institute’s composite rating in simple terms as being 10%. This might seem a lot. It isn’t. The most accurate measurements ever used to gauge qualities like student achievement and ability often fail to realize such a low level of unreliability. It’s very challenging to measure knowledge and skill with great precision.

A basic formula from the mathematical theory of measurement can be used to estimate how unreliable this particular rating is. Theoretically, suppose we could measure each part that contributes to the Institute’s composite rating 20 times this year for every student. For each of these 20 sets of measurements, we compute a new rating, 20 hypothetical ratings in all. This provides a statistical scenario for estimating how much error there is in the Institute’s rating of a school.

Though the Institute rated the average elementary school at 6.2, we know there is error in this number. The theory tells how much. Nineteen times out of 20, this school’s actual rating could be as high as 7.3 or as low as 5.1. Odds are that one of the 20 ratings would be outside this range. To put it roughly, any school’s rating might be about 1.1 points lower or 1.1 points higher than the rating published by the Institute.

Now, imagine a new professor who is joining the faculty at SFU. Which elementary school in the area surrounding the campus—Burnaby, Coquitlam, and New Westminister—would this professor’s family choose based only on the Institute’s ratings?

The 104 schools in these three districts have quite a spread of Institute ratings, from 3.1 all the way to 10.0. The family might think, “Let’s go for the top rated school!” But the mathematical theory of measurement reveals some findings that may interest them.

Because the Institute’s ratings are not perfectly reliable, only schools with ratings lower than 7.8 can be distinguished from the top-rated school 19 times out of 20. Thus, there are 29 top-rated schools from which the family can choose. Or, suppose the family is really interested to choose a school that’s above average. This turns out to be quite challenging. Taking into account unreliability of the Institute’s ratings, 19 times out of 20 only three of 104 schools across these three districts can be reliably considered above average. Only seven schools can be considered below average.

Our advice for parents and others who use the Institute’s or anyone else’s ratings is plain. Don’t place too much emphasis on any single indicator when judging a school because some differences are probably just due to unreliability. Exam scores have a role in making judgments about education. But limitations like unreliability should be taken into account. When this is

done, it becomes clear that parents can choose among a wide range of schools on the basis of other important factors, such as how far from home it is or whether it has programs that match their children's special interests.

Are Other Indicators Worth Considering?

Exam scores aren't the only indicators of students' learning, teaching quality or a school's effectiveness. For example, many readers are parents of grade 12 students (or are students themselves!) who are applying for admission to post-secondary education. In BC, a number of course grades on these students' transcripts are another composite indicator. It's made up of 60% of the teacher's grade on classroom work and 40% of a provincial exam score.

To see the value of combining measures from different sources, consider this situation. Suppose you knew a student's provincial exam scores on English 12 and on Principles of Mathematics 12. How well could you predict the teachers' grades for this student in these two subjects?

Take a student who scored exactly at the average on the provincial exams. That's 70% or a C+ on the English 12 exam, and 72% or a C+ on the Math 12 exam. What range of teachers' grades would be predicted for this student?

We used basic statistics to tell us what range to expect when measuring this student's accomplishments, say, 20 times. For 19 of those times, a formula that uses the statistical correlation between provincial exam scores and teachers' grades indicates the teacher's grade might be as low as 49% in English. This corresponds to a letter grade of F. On the bright side, the teacher's grade might be as high as 92%. That's a solid A. In Math, the predicted teacher's grade has approximately the same range. On the low end it's 52% which is just a C-. On the high end, the teacher's grade is predicted to be 91%, another solid A. These are large ranges!

Maybe you want to start with teacher's grades and predict provincial exam scores? The results are nearly identical.

The point is that standardized exams, whether provincials or FSA, measure different things than teachers' grades. So, which should be used? Exams? Teacher's grades? If you follow our line of thinking, you may have guessed our answer. It's both. Provincial and FSA exams measure a very small sample of what students might learn in English or Math but they do so in a very controlled way. Teachers' grades reflect a much broader sample of knowledge and skills taught in these courses and may take advantage of a wider variety of techniques for measuring these accomplishments. At the same time, technical strengths of teachers' grades differ from those of the provincial exams.

What's the Upshot?

Choosing a school, assessing teacher effectiveness and evaluating school quality are important and complex issues for BC's citizens and its government. People need high quality information to tackle these issues.

Reaching informed positions on these issues is not simple. Good intentions, good reasoning and open-mindedness are necessary but may not be enough. Technical knowledge and statistical tools are useful, too. First, they help identify problems that can arise when judging whether goals are met. Second, they can alert us to important issues when trying to rate whether one school is better than another. We hope our discussion of the Fraser Institute's ratings help people make better decisions about schools.

We've tried to demonstrate why it's important to heed these cautions. It's because any single indicator, composite or otherwise, almost always misses the mark to some degree. Some indicators may actually address different topics than you might think at first.

What about using the Fraser Institute's ratings? If people use them, we urge they first take time to develop a thorough understanding about how this indicator is constructed and what its limitations are. We add quickly that the same applies to proposals we make about how to interpret ratings and ratings of any sort.

Education is critically important to the well being of children, society and the economy. We urge everyone to seek answers to important questions about this vital undertaking. When our children and our collective futures are at stake, we can't afford to behave like the novice carpenter: Measure once, cut twice. Once opportunities for students' and schools' are cut, changing that decision can be very difficult.