

Comment on Bishop: Are national exit examinations important for educational efficiency?

Anders Björklund*

This is an important and impressive paper. It is important because it addresses a concrete policy issue, i.e., whether national exit examinations serve a useful purpose in a nation's school policy. And because the analysis in the paper is relevant for all kinds of national examinations of schools and of pupils, not only examinations at the time of exit of, e.g., high school. It is impressive because it is based on the author's collection of a lot of data from many different sources.

Obviously, we would like to get empirical evidence on this issue: Is there an impact—a causal effect—of national examinations on the achievement of students? John Bishop has found that there are two sources of variation in data generated “in the real world” that can be useful to inferring effects about such examinations. First, there is *cross-national variation* in the use of such examinations: some countries have such tests and some do not have them. Second, there is *cross-regional variation* within Canada in the use of such examinations. Both the cross-national and cross-regional analyses suggest that national examinations promote the achievement of students.

John Bishop also mentions an alternative way of assessing the impact of national examinations, namely, to study the effects of eliminating them. A short analysis of what happened in Sweden in the 1970s, when national exit examinations of Swedish high schools were eliminated, basically confirms the results from the cross-national and cross-regional analyses.

I think that the most promising way of getting (even) more convincing empirical evidence on the impact of national examinations is to use variation over time in a country. For Sweden, it might be possible to exploit the frequent changes in the use of national examinations (*centrala prov*, *standardprov* and *nationella prov*) for further analysis

* Professor at the Swedish Institute for Social Research, Stockholm University.

of these issues. If national examinations have an effect, we would expect to see improvements in those subjects where the tests have been introduced. If there have been general changes in the achievement of students, we would expect to see a more favourable change over time—higher increase or lower decrease—than in subjects for which examinations have not been introduced. The opposite pattern would be expected in subjects in which national examinations have been eliminated.

A further requirement for such a study would be information from independent tests that tell us about skills in relevant subjects before and after the changes in national examination policy. For example, the tests of numerical and verbal skills given by the Education department among (random samples of) 13-year-olds since 1961, would have been ideal to use if national tests had been introduced (or eliminated) in Swedish or Math at this stage of the school system.¹ Stimulated by Bishop's paper, I made some attempts to find out whether these two informational requirements—changes in national examinations and information from tests of the relevant skills—are fulfilled in any part of the Swedish school system. I did not find any such case, but maybe someone who knows more about the history of Swedish school examinations could find one.

An alternative way of learning more about the impact of national examinations would be to impose such examinations in a controlled—possibly randomised—experiment. The Swedish *Skolverket* could randomly select several municipalities to be the experimental group in a study of the impact of national examinations. These municipalities are told before the school year starts that a national test will be done in all schools in the municipality at the end of the school year to examine whether the students have learnt what they are supposed to learn. The local politicians will be informed about the results of the test. In another group of municipalities—the control group—the same test will be done at the end of the school year, but neither the schools nor the local politicians will be informed about the test until quite late.² If the experimental and control groups of municipalities are made “reasonably” large, the impact of the national examination can be estimated as the difference in outcome between the two groups.

¹ See, e.g., Emanuelsson et al. (1993).

² This test would not be necessary if there is information from an alternative test given in both experimental and control municipalities.

In Sweden we are not used to such randomised experiments, so it might be difficult to get the political support that is needed. The now famous STAR experiment in the state of Tennessee, where the impact of smaller class sizes was examined in a large-scale randomised experiment, shows that quite sophisticated randomised experiments are feasible. See Krueger (1999). Some day this approach to evaluation of public policies in the fields of education and labour market policy must come to Sweden too.

References

- Emanuelsson, I., S-E. Reuterberg and A. Svensson, (1993), Changing Differences in Intelligence? Comparisons Between Groups of 13-Year-Olds Tested From 1960 to 1990, *Scandinavian Journal of Educational Research* 37, 259-277.
- Krueger, A. B. (1999), Experimental Estimates of Educational Production Functions, *The Quarterly Journal of Economics* CXIV, 497-532.

