

Progress in mathematics during earlier years in Swedish school

Lars Adiels
Ph.D. and Ph.D. student UTEP,SU

Abstract

The results of the improvement in Math between school year 3-4, 5-6-7 and 8-9 in the Swedish school system is analysed using the effect-size estimator. The result shows that the yearly improvement decreases in particular when the pupils reach school year 7-9. The estimate is based on the Swedish version of the international kangaroo competition. A few points on reliability are discussed. The validity of using this particular data is also discussed.

Introduction

In an ongoing research project at Stockholm University we try to establish an environment within ordinary school which is able to evaluate quality of the education and whether systematic changes in pedagogy can improve the quality. This work will start with mathematics at school year 10 in the Swedish school system. The design of the project makes a clean setup for using the effect size estimator. By a formative test at the beginning of the course and than virtually the same test at the end of the course, one can use the effect size estimator as a quality estimator, or to test new pedagogical approaches in order to establish better practice. In the later case measurement will be applied to a bigger group of classes that serves as a control group and to a smaller group. In the smaller group some, hypothetical better, systematic approach is applied. The effect size is measured and compared.

A worry is commonly expressed that students in Sweden are loosing ground in Mathematics. This is expressed inparticular by politicians analysing PISA [SR (2008)] and other international tests. As a proof of concept the method has been applied to publicly available empirical data and the progress in math for children age 9-15 was estimated. The results show a drastic drop in performance of Swedish math students at the age of 13-15. This finding may have implications when the curriculum for the Swedish school system is to be thoroughly revised.

Granted by TUFF research school ...

The effect-size estimator

The estimator effect-size is basically the difference between the performances at two distinct assessments divided by the standard deviation. This estimator is used to make Meta-analyzes where it is not so clear which standard deviation one can measure. There are therefore different definitions of effect size Cohen's d[Cohen (1988)], Glass's Δ[Hedges & Olkin (1985)] and Hedges' g[Hedges (1981)]. The main difference is how to calculate the standard deviation. The basic definition is using the pooled standard deviation but sometimes this is not easy calculable. There are arguments that say that different definitions of the effect size should give similar results, as in Cohen's original proposal, if an experiment or Meta analyses is well performed. [Coe (2004),Tymms (2004)p56]

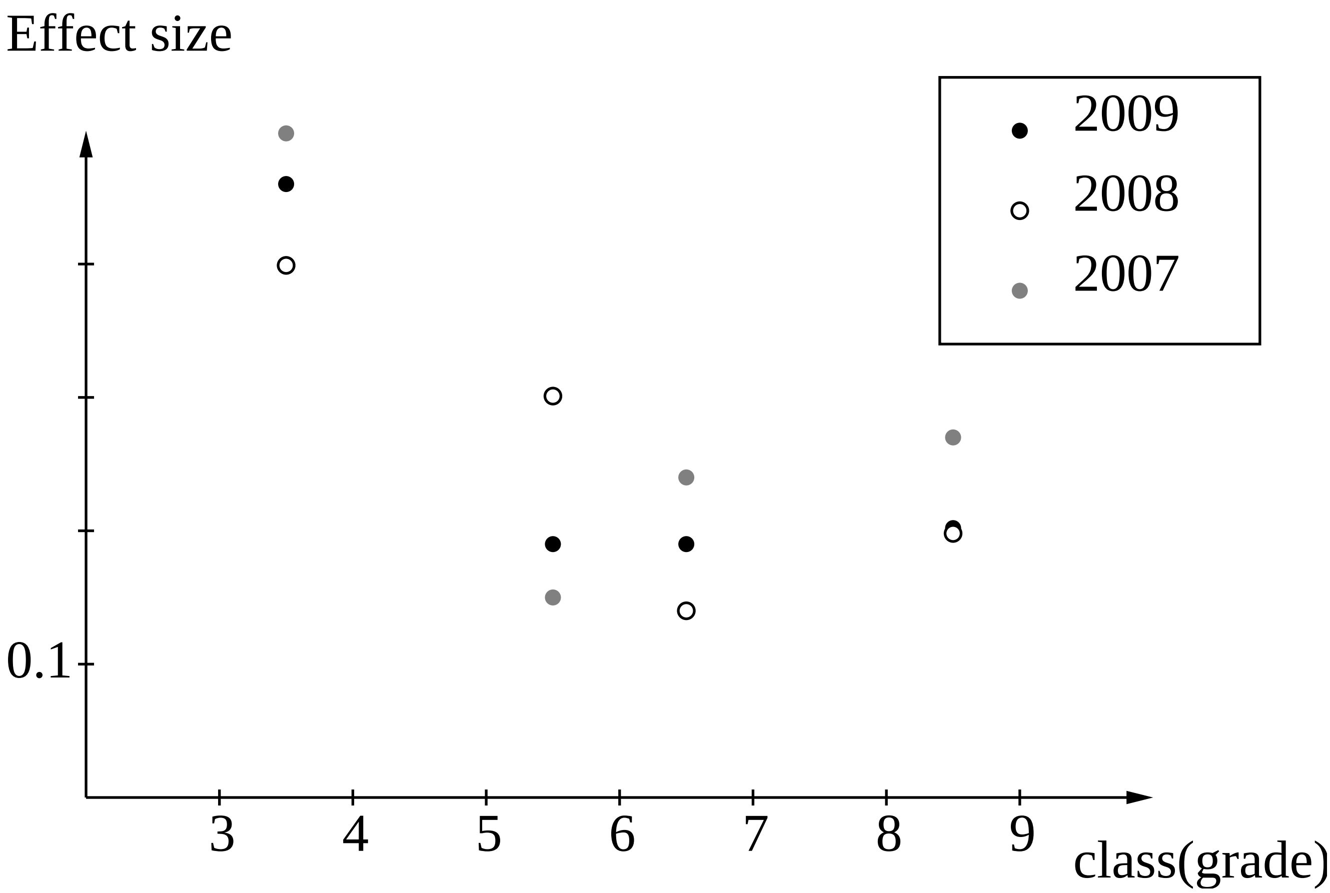
Cohen claimed that effect sizes smaller than .2 are considered very small, when in the range .2-.5 are considered small but notable; .5-.8 are considered medium and above .8 are considered large. Several international Meta studies have been preformed using the effect-size estimator. Hattie [Hattie (2008)] and others express that most students in average should have an effect size of 1 in three years. He also claims that a student loses a fraction of "gained" effect size during summer vacation. He compares different pedagogical approaches and is therefore considering approaches that give an effect size above .4 per/year worthwhile and approaches that give below .4 per/year as bad. An other way to use effect size could be to use it as a quality measure. If we run the same formative test every year one could evaluate if quality of education is kept, increased or decreased over years.

There are also criticisms on the use of effect size measures. The effect size estimator requires that we work with normal distributed data. This is not fully in line with modern test where you try to find knowledge rather then discriminate between students. However Dylan [Dylan (2007)] shows how within normal groups of students there is a normal distribution with a variation of 1 standard deviation which corresponds to three school years. This suggests that if a test is properly inline with the course there are good possibilities that the result will be normal distributed. The name effect size suggests causality which is not necessarily true. [Godfrey (2004)]. However effect size is used within educational research and can be utilized for comparison if we have a similar setup between experiments.

The analysed data samples

As a proof of concept some publicly available real data samples were studied. The data is taken from the international math competion Kangaroo. The kangaroo competition in Sweden is not so much of a competion as it is a way to stimulate an interesting discussion for all students. However it starts up as a competition and all results from classes are reported and collected. On each year there are cohorts between 2000-8000 students from the different levels and different years (2007-2009). These represent mainly full classes or groups. The interesting thing with the kangaroo tests is that the very same test is used for two or three grades. Thus it is possible to compare grade 3-4, 5-6-7 and 8-9. Those grades are all having the same curriculum over the country. One could also compare at upper secondary level but there not all students have the same curriculum.

The standard deviations for the single samples were calculated. The standard deviations varied with about 1 unit for grade 3-4 and 8-9 except for one case and with about two units for the grade 5-6-7. The minimum standard deviation is 11.5 and the highest



Figur 1. The measured effect size 2007-2009, between age groups (classes)

is about 17. The pooled standard deviations are of course between the individual values. Average is varying as progress is supposed to occur between different grades. The results are summarized in table 1. In figure one the data is drawn as a diagram. For one sample there seems to be an irregularity but I used the raw data and did not try to correct it in any way. Also sample sizes are given as published. In the table grade 6 is given twice as it can be compared with both grade 5 and grade 7. This data (2007-2009) can be found at http://ncm.gu.se/kanguru and following pages. The work was done in order to mimic how data could behave in future experiments, but an interesting finding of general interest was noticed. There is a decrease in effect size at grades 7-9, or perhaps even earlier. This finding will be discussed in the next section.

Findings and discussion

The data shows a decrease in effect size for years 6-9 or even from 5^{th} grade. This gives an indication that math problems starts later then after the first years in school. There are many possible objections against this observation. Some of the objections will be discussed here.

- The kangaroo math is not relevant to study math development at all. But a lot of very experienced teachers and researchers have put a lot of effort into those tests. They are there to inspire also students that are not so strong in conventional school math. They represent some international consensus on what is good for school children's math development.

- The validity of the data. This can be subdivided in many different questions and some will be discussed one by one.
- Different sample sizes. In particular between grade 3 and 4, and to a lesser extent between grade 5 and 6, there is a substantial difference between the sample sizes. The cause for this is unknown. If this effect comes from that only interested teachers (or teachers that believe that they have good students at low age or at a new school in grade 7) report the results of the test, then this should rather point in the opposite direction and reduce the observed effect.
- Not all teachers report all students. To study this effect one should go back to original raw data which has not been possible at this occasion. If one looks at the distributions by naked eye one observes a shoulder compared with a bell curve towards higher scores. This increases the standard deviation and reduces effect size if true, thus underestimating the effect size where the effect is strongest. On the other hand the curves for grade 6 and 7 almost overlap except for the tails for all three years indicating an even smaller improvement except for a small fraction at the top and bottom. That would mean that the students improved even less.
- Effect size requires normal distribution. It is clearly not true in all cases. There are other estimators that are not so much affected from errors in the tails. Further, this is a multiple choice test. It should be "impossible" to have less than 20% with five possible answers. Of course some students don't answer all questions due to lack of time. Still there are some obscurities in the data in this respect. Those effects could possibly be dealt with if one go back to the original data.
- There is no international comparison. The way data is displayed to public is different between countries. Some country focuses on competition whereas some other countries like Sweden focuses on widening the interest in Math.
- For later years (grade 10-12 not shown here) data is biased by choice, students that like math choose it. The increase in effect size is also huge.

Conclusion

Data suggests that the math development in Swedish school, in terms of effect size, decreases somewhere around the age of 12. There is the interesting possibility to extend this investigation to international kanguroo test and to perform an international comparison.

Table 1

class	Year				Year				Year			
	2009				2008				2007			
	SD (no pool)	average	effect	sample	SD (no pool)	average	effect	sample	SD (no pool)	average	effect	sample
3	11,65	27,5	0,48	2842	13,72	34,5	0,4	5241	12,19	27,2	0,5	3308
4	12,7	33,1	0,44	5854	14,78	40,0	0,37	7763	13,02	33,3	0,47	5690
5	11,77	26,9	0,2	3467	14,0	29,8	0,32	3736	14,52	32,6	0,16	2904
6	12,82	29,2	0,18	4457	15,2	34,2	0,3	6253	17,15	35	0,14	5713
6	12,82	29,2	0,19	4457	15,2	34,2	0,14	6253	17,15	35	0,24	5713
7	13,51	31,7	0,18	3396	15,9	36,3	0,13	5576	16,34	39,1	0,25	4360
8	11,97	27,8	0,22	2368	12,6	25,4	0,2	2638	12,21	29,4	0,28	2368
9	12,46	30,4	0,21	1934	14,44	27,9	0,18	2051	13,5	32,8	0,25	2051

References

Coe, R. (2004). Issues arising from the use of effect sizes in analysing and reporting research. In I. Elliot K. Shagen (Ed.), *But what does it mean? the use of effect sizes in educational research.* (p. 90). Slough: NFER.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2. ed. ed.). Hillsdale: L. Erlbaum Associates.

Dylan, W. (2007). *Seminar for a new PhD resarch school in assessment.*

Godfrey, . (2004). Effect size: a statistician's pseudo-concept? In I. Elliot K. Shagen (Ed.), *But what does it mean? the use of effect sizes in educational research.* (pp. 101 – 109). Slough: NFER.

Hattie, J. (2008). *Developing potentials for learning: Evidence, assessment, and progress.* Stockholm. Available from `http://www.did.su.se/content/1/c6/04/73/80/JohnHattie.pdf#.#`

Hedges, L. V. (1981). Distribution Theory for Glass's Estimator of Effect size and Related Estimators. *Journal of Educational and Behavioral Statistics*, *6*(2), 107-128. Available from `http://jeb.sagepub.com/cgi/content/abstract/6/2/107`

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Orlando: Academic Press.

SR. (2008). Program series in Swedish Radio. Available from `http://www.sr.se/sida/artikel.aspx?ProgramID=3238&Artikel=2250145`

Tymms, P. (2004). Effect sizes in multiievel models. In I. Elliot K. Shagen (Ed.), *But what does it mean? the use of effect sizes in educational research.* (p. 56). Slough: NFER.