Pre-print of article published in *Child: care, health and development, 2009,* 35(6), 858-867

An alternative design for small-scale health experiments in schools: Does daily walking produce benefits in physical performance?

Abstract

Background

Low levels of physical activity in childhood may cause reduced physical performance. The aim of the study was to assess health benefits of twenty minutes daily walking during one school year for children of age 6-14, and to utilize an alternative method of analyses.

Methods

In a rural school in inland Norway all school children joined the project. Measurements of low back static endurance, hamstrings flexibility, standing balance and cardio-vascular fitness were made before and after the intervention.

Intervention and “aging” are confounded. A statistical model is adapted to resolve this issue without the need of an independent control group.

Results

Age-adjusted analyses showed 11 % increase in low back static endurance, 8 % increase in hamstrings flexibility, 69 % increase in balance, and 6-13 % increase in cardio-vascular fitness. The effects were largest among those children who had the lowest performances before the intervention.

Key words: Small-scale school experimental design, Physical performance, walking, school children.

BACKGROUND

As our lifestyle becomes increasingly passive in the industrialised world, there are some indications that physical health of young people is declining, although the literature is sparse on the subject. There are theories that the steadily increasing amounts of sitting in daily life may constitute a risk for poor physical health, including a variety of disorders (Pate et. al, 1995, Beaglehole & Bonita, 1999). Results from the cross-cultural Kraus-Weber-investigation in the 1950s indicated a reduction over time in trunk strength and mobility among school children (Kraus & Raab, 1961). Findings in a recent Swedish study comparing 16-year old adolescents in 1974 and 1995 indicated a decrease in hip mobility over time (Westerstahl, 2003). Several recent studies indicate a decrease in cardio-respiratory fitness among children and adolescents from 1980 till 2000 (Tomkinson, 2003). A large Norwegian study comparing physical performance of 15-year old adolescents in 1968 and 1997 concluded that low back extension endurance and ankle mobility were poorer in 1997 than in 1968 (Sjolie & Mønness, 2007**.** Cross-cultural differences in physical performance have been found between European and American school children (Kraus & Hirschland, 1954), and between Arabs and Chinese adults versus Caucasians (Ahlberg et. al, 1988, Hoaglund et. al, 1973).

Walking to school is a mode of physical activity that has been in decline from 1985 till 1992 according to British and Norwegian surveys (Department of Transport, 1995, Vibe, 1993). Some studies indicate that use of school bus is associated with low levels of physical performance, and that regular walking is associated with higher levels of physical performance (Kraus & Raab, 1961, Kraus & Hirschland, 1954, Solstad, 1973, Sjolie, 2000).

The headmaster at a small Norwegian rural school, where most children used school bus, was worried because of an impression of a decline in physical performance in the pupils over time, and wanted by simple means to improve their physical health.

AIMS

The aim is two-fold:

To present a design for small-scale school experiments, and to present a study to explore if short daily walking trips, performed for one year, would produce health effects in primary school children. The hypotheses were that walking would increase low back extension endurance, balance, hip mobility and cardio-vascular fitness.

METHODS

Material

The study population included all the 105 pupils in January 2001 at Plassen primary school, which is one of the two primary schools in the Alvdal municipality of Hedmark County. The children's age varied from six to fourteen years in school year one to seven, respectively. Approximately three quarters of the children were bussed to school, due to long distances and a heavy traffic on the surrounding roads. However, the bussing pattern was varied, making it impossible to incorporate this effect in the project. A green area of forest and a river surrounds the school, with good opportunities for hiking. The headmaster insisted that all children at the school should join the project, and gave ethical and practical reasons for that decision. Often in school research settings, "*political, practical and ethical constraints have limited the use of randomization*" (Bryk & Weisberg, 1976). Due to this objection the idea of dividing the classes to make intervention and control groups was abandoned. The school council was informed about the study and gave their consent in spring 2000. All parents and children were informed by several information letters and in several meetings about the project during the year 2000. As the study was initiated and conducted by the school, passive informed consent from the children and their parents was sufficient, according to Norwegian law at that time. One child informed in a letter, signed by the parents and her self, that she did not want to be measured. Another child was excluded due to a hip fracture in December 2000 that resulted in wheel chair use at the start of the project. One hundred and three pupils were measured in January 2001 a few days before the project started, and 85 were re-measured in January 2002. Three pupils had moved to other municipalities, and all pupils in school year seven had moved to the junior secondary school in the municipality. Thus the data consist of 100 measurements before the project started and 85 measurements afterwards.

Intervention

The intervention consisted of 20 minutes daily walking during school hour. Different rugged forest trails, 1.5-2 km long were used, providing opportunities for a variety of gradients, steepness, climbing and balancing. The pupils walked in class groups, led by their teacher. The walk was performed for every school day from 2.2.2001 until 2.2.2002, except for three days when the walking was substituted by indoor physical activity because the weather was too cold (-30ºC or below). All pupils joined the walking all days; no one ever tried to withdraw or had letters from the parents claiming non-participation any time.

Statistical methods

Each property is measured on each pupil at the start of the experiment, and one year later, as a repeated measurement design.

Let Y_{cji} be the property in question, measured at school year c, $c=1...7$, on pupil j, $j=1...100$. The index $i=0$ indicate "before intervention", and $i=1$ indicate "after intervention". When $i=1$, the index j only runs to 85. The measurements have the following pattern (Table 1):

There are of course no measurements after intervention corresponding to school year 1, and we do not know how the pupils in school year 7 measured before the intervention performed a year later, as they have left the school. The individual effect on pupil j of the intervention could be measured by $\left(Y_{cj1} - Y_{c-1j0}\right)$. This figure measures the change of performance, from before the intervention to after the intervention. However, the pupil has also grown in age. Intervention and “age increase” are confounded. Figure 1 shows the measurements of the balance property (se the physical measure chapter later). We see clearly that there is an age effect in addition to an intervention effect. The effect of intervention might be strongly biased if ignoring the effect of aging.

Since body control and strength are in major development in these age groups this issue must be addressed. Ideally, we should have known the pupil’s potential performance if not undergone the intervention, which is impossible, of course. We need an estimate of the effect of aging to get an unbiased estimate of intervention effect.

The usual procedure is to establish an independent control group. However, establishing a control group by splitting each class was as mentioned considered to be out of question by the headmaster due to ethical and practical reasons. Unethical, because the headmaster did not want to divide the class and exclude some children from the walking. Impractical, since any control group probably might have seen it as a competition and started to practice for themselves. This kind of experiment can not be “blind”; everything is open to be seen and possibly copied by any other person. An alternative is to involve another school. This only partially solves the no-blindness issue, doubles the cost and increases project management. Also, another school may have other systematic differences (there might be a “school effect”), as discussed later.

However, it is possible to adjust for an age effect by utilization the data as explained: The development of the “Before intervention” data, from school year to school year, can yield an estimate of the average effect of ageing, with no intervention effect involved. The figure

$\left(\overline{Y_{c\bullet 0}} - \overline{Y_{c-1\bullet 0}}\right)$ estimates the <u>average</u> ageing effect. ($\overline{Y_{c\bullet 0}}$ is the mean of the pupils in school year *c* before intervention.)

This estimator can thus be subtracted to allow for a measure of intervention effect that is not confounded with the average age effect.

Let $\Delta Y_{cj} = \left(Y_{cj1} - Y_{c-1j0}\right) - \left(\overline{Y_{c\bullet 0}} - \overline{Y_{c-1\bullet 0}}\right)$, *c=2…7* and *j=1…85*.

(There are only 85 “after intervention” measurements).

If the intervention does not have any effect the expectation of ΔY_{cj} will be zero.

Between school years, the ΔY_{cj} will only be statistically dependent between adjacent school years. Within a school year, there is dependence since $\left(\overline{Y_{c\bullet 0}} - \overline{Y_{c-1\bullet 0}}\right)$ is constant for all pupils within the school year, but this does not alter the individual differences. It is possible but tedious and extensive to adjust for the dependence when calculating the significance levels and we believe the statistical correction to be of minimal effect. Only the standard errors of the estimated effects are slightly underestimated, not the estimated effects themselves. Thus the significances shown are calculated as if the ΔY_{cj}s are statistically independent. We will use ΔY_{cj} as the dependent variable in a linear model with intervention, school year, gender and the “individual performance before intervention level” (Y_{cj0}) as the independent variables.

All calculations are performed by ordinary least squares regression, using SYSTAT (SYSTAT, 1998).

Physical measurements

The procedure of the measurements followed a written manual, which described in detail the measurements. The tests were performed in the same sequence between 9am and 2pm in two adjoining rooms at the school, where the temperature was kept at 21^{o}C. The same equipment was used for all children, and was identical at both occasions except for a pulse measure. None of the pupils had joined physical education classes or swimming the day they were examined. The tests of the low back, hip mobility and standing balance was performed by a trained physiotherapist, while the fitness test was done by a teacher of physical education.

Low back extension endurance was measured as static endurance in accordance with Alaranta´s modification of Sorensen´s test (Alaranta, 1994). From a prone forward bent position, with the iliac crest on the edge of the couch while stabilized by the ankles with one strap, the children elevated the upper half of their bodies to the horizontal plane, as marked by a measure stick. The elevated position of the trunk should be kept as long as possible up to 4 minutes. Endurance time was measured in seconds by a stopwatch.

Hip mobility was measured as hamstrings flexibility by Active Knee Extension Test (Gajdosik & Lusin, 1983) with a prolonged steel goniometer which displayed singular degrees. The right hip was measured, or the left hip in case of minor traumas on the right side that could influence the performance. The knee was stabilized with a locked Don-Joy–orthosis, and the lumbar region was stabilized with the pupils' own hands in the back.

Standing balance was measured as time spent standing fixed on one foot up to one minute, while the eyes were closed (Byl & Sinnott, 1984). Standing time was measured in seconds by a stopwatch.

Cardio-vascular fitness was measured by the step-test (Watkins, 1984). The heart rate was measured electronically after the pupils had ceased walking up and down on a 25 cm bench for four min, keeping a steady speed in accordance with a metronome set at 120 beats per min. The heart rate was measured. The heart rate was measured immediately after the performance, one minute afterwards, and two min afterwards. The measure device was Polar Accurex + in 2001, and Timex Ironman in 2002, because the Polar device was out of function in January 2002.

The intra-tester reproducibility of the testing physiotherapist was previously assessed by two measurements on 11 healthy persons, made at two weeks interval. The correlations (Pearson's **R**) between the tests of low back endurance, hamstrings flexibility and balance were 0.98, 0.89 and 0.95, respectively, and the corresponding coefficients of variance (standard deviation of difference divided by the mean of the means) were 4.8 %, 9.6 % and 19.8 %, respectively. The correspondence between the two heart rate measure devices was tested by simultaneous measurements on eight pupils immediately after they had performed the stepping for four min, and repeated one min afterwards. The correlation was 1.0 for both occasions, and the corresponding coefficients of variance were 1.8 % and 3.5 %.

RESULTS

Mean values before and after the intervention in total and for the different school years are displayed in Table 2.

There was a clear significant intervention effect on all measured properties (Table 3).

Balance was improved by 69 %, heart rate one and two minutes after the step test was reduced by about 14 %, and low back endurance was increased by 11 %. However, the decline in heart rate was not significant with this model. Table 3 also shows % performance increase if not corrected for an age effect. Generally these values are higher, and (not shown) with seemingly better significances. The difference is most evident with balance and low back endurance. Inspecting table 2, we se that those properties have the most evident "age effect" = monotone growth of the before intervention data.

The next hypothesis is that the intervention effect also varies with school year, gender and the individual property level before the intervention (Table 4).

The intervention effect on balance was not affected by school year. Else the intervention effect was depended by school year and performance level before the intervention, giving higher effect for those displaying lower values of physical performance initially (The regression coefficient of Y_{cj0} is negative). Gender was of minor influence. If analysing uncorrected data, only tests for overall intervention and school year would be affected. Tests for gender and performance level before intervention must be identical, since the adjustments align with the school year effect. Figure 4 depicts the balance property. It confirms table 4; there is an intervention effect (most points are above zero), there is hardly any year or gender effect. Most girls in school year 7 performed poorly.

DISCUSSION

Within school research, observational studies are more common than randomised clinical trials, as in medicine (Bryk & Weisberg, 1976). There is a huge literature on how to measure effect of an intervention without having a control group. The idea is using growth curves: Each child is supposed to have a natural time depended development. An intervention may then be a value-added effect. In order to measure the added effect, some estimate of the natural growth has to be found and subtracted. See Heck (2000), Thum (2003) Seltzer et. al. (2003), Lockwood et. al. (2007) among others. Here we use a simple estimate of natural improvement, comparable to having a control group.

Most pros and cons of clinical trials certainly prevail here too. We will compare with the general setup of clinical trials with a treatment group and a control group. Analysing *outcome* or *change from baseline* is discussed in Altman (1991). Here we follow the outline found in

Everitt (2004). Note that our alternative analysis only works since we have the special "school year structure" that yield data that mimic having a control group at each school year.

Let X_{cji} be the measure from a hypothetical independent control group, measured at school year *c, c=1…7*, on pupil *j, j=1…J.* The index *i=0* indicate "before intervention", and *i=1* indicate "after intervention". Everitt discuss three methods; POST (outcome), CHANGE (from baseline) and ANCOVA.

POST

POST is just to ignore the before intervention data and estimate *Year*intervention* effect by $\left(\overline{Y_{c\bullet 1}} - \overline{X_{c\bullet 1}}\right)$. In our case, this could be mimicked without a control group by using $\left(\overline{Y_{c\bullet 1}} - \overline{Y_{c\bullet 0}}\right)$ as an estimator of *Year*Intervention* effect. However, this should not be called a POST since we actually compare before and after intervention data, between two different pupil groups but at the same school year level. An individual response of the intervention is also available.

CHANGE

With an independent control group, $\delta Y_{cj} = \left(Y_{cj1} - Y_{c-1j0}\right) - \left(\overline{X_{c\bullet 1}} - \overline{X_{c-1\bullet 0}}\right)$ would have been the individual CHANGE response estimate. We introduced $\Delta Y_{cj} = \left(Y_{cj1} - Y_{c-1j0}\right) - \left(\overline{Y_{c\bullet 0}} - \overline{Y_{c-1\bullet 0}}\right)$ as the individual intervention response. Comparing to Bryk (1976), $\left(\overline{Y_{c\bullet 0}} - \overline{Y_{c-1\bullet 0}}\right)$ act as the natural growth. Some issues of statistical dependence discussed with ΔY_{cj} also concerns δY_{cj}. Taking means give estimates of *Year*Intervention* effect. Note that $\overline{\Delta Y_{c\bullet}} = \left(\overline{Y_{c\bullet 1}} - \overline{Y_{c\bullet 0}}\right)$ is the same estimator as shown at POST since factors cancel out when taking means.

ANCOVA

Reshuffling within the brackets yields

$$\delta Y_{cj} = \left(Y_{cj1} - Y_{c-1j0}\right) - \left(\overline{X_{c\bullet 1}} - \overline{X_{c-1\bullet 0}}\right) = \left(Y_{cj1} - \overline{X_{c\bullet 1}}\right) - \left(Y_{c-1j0} - \overline{X_{c-1\bullet 0}}\right)$$

With this last formulation, δY_{cj} can be seen as how much better pupil *j* is compared to the control group after intervention, minus how much better he was before the intervention. Due to the effect of *regression to the mean* (Everitt, 2004), the correction could be to strong, so it is advised to estimate the correction size, introducing the response estimate

$$\delta' Y_{cj} = \left(Y_{cj1} - \overline{X_{c\bullet 1}}\right) - \beta\left(Y_{c-1j0} - \overline{X_{c-1\bullet 0}}\right).$$

β could be estimated from the regression

After-intervention-data = constant + β*Before-intervention-data, preferably from the control group. ANCOVA includes POST ($\beta=0$) and CHANGE ($\beta=1$).

This could equally have been done in our case, reshuffling within the brackets and introducing a β gives

$$\Delta' Y_{cj} = \left(Y_{cj1} - \overline{Y_{c\bullet 0}}\right) - \beta\left(Y_{c-1j0} - \overline{Y_{c-1\bullet 0}}\right).$$

Estimation of a β in the same manner is in this case more dubious since the intervention could influence the estimate. Note that $\overline{\Delta' Y_{cj}} = \left(\overline{Y_{c\bullet 1}} - \overline{Y_{c\bullet 0}}\right)$ is again the *Year*Intervention* estimate since the mean of the second bracket is zero. , Thus the estimate is independent of β; the value of β would only influence the analysis of individual intervention responses.

The estimated βs in our data (not shown) are between 0.4 and 0.7, thus the correction on individual level would have been slightly smaller than the one we have used in the actual data analysis shown.

An alternative to $\left(\overline{Y_{c\bullet 0}} - \overline{Y_{c-1\bullet 0}}\right)$ as an age correction could be to implement a parametric growth curve based on the pre-treatment data (Bryk, 1976).

During the experiment, influencing effects outside the experiment might have changed. There might be a “society change during a year”. The pupils may change attitude to physical activity in general. This can not be addressed without an independent control group, but also a control group could change their attitude. However, having another school as a control group could introduce a “school effect” bias. Maybe the other school has a different (in this case physical) attitude? Our model avoids this. Since the ideal case (having measures with and without intervention on the same pupil) is impossible there will be a trade-off between competing issues; the headmaster’s ethical issues, the practicalities and increase of variation (=bias) by involving more schools and the “society change during a year”. In this case we believe that “society change during a year” is of minor influence, so that conclusions based on ΔY_{cj} should be perfectly acceptable. The statistical method is especially valuable in a school setting were there are both successive age groups and age dependent interventions and could be of value in many small-scale school experimental situations. In a large-scale setting, an independent control group is certainly preferable.

Independent of above model choice, other information, like what the pupils do in leisure time, family information etc. could have been adjusted for if available and allowed to be used by the school and the parents.

The main findings in this study were that physical activity in the mode of daily walking for twenty minutes during one school year increased significantly low back extension endurance, hamstrings flexibility, standing balance and cardio-vascular fitness in children. The health effects were strongest among those children who had the poorest physical performance before the intervention. All statistical significances are from a two sided test. Since a negative effect (in body sense) could be disregarded, the significance is even stronger. As the sample is small the results must be interpreted with caution. We have tried to adjust the data for the age effect on body development. Even if this introduces dependence in the data, we believe this is better than just ignoring this effect, as one would do applying a plain repeated measurement design. The adjustment is based on the assumption that there is some kind of monotone ageing trend in body strength and control. However, the data from "school year 2 before intervention" shows that this group have the highest hamstring value, a high back endurance, and also the lowest pulse rate just after the step test. All these values are better than expected if one anticipate a monotone age trend. Also the girls entering "school year 7 after the intervention" had worse balance results than they did in "school year 6 before intervention" The adjusting method is vulnerable to such odd pattern, see figure 2.

The reliability and validity of the test of low back extension endurance was tested previously among adolescents (Oksanen & Salminen, 1996, Sjolie, 2000). An identical test of hamstrings flexibility among school children showed good reliability and validity (Pettersen, 1997, Sjolie, 2000) and Active Knee Extension Test was also tested for reliability and validity among adults (Gajdosik & Lusin, 1983). The balance test had the poorest reliability in our test of adolescents. The test was also easy to remember and to perform, and the pupils may have practiced the test for themselves. These factors may have influenced the high effect value on balance. The test has been used for school children (Johnsson, 1983) and adults (Byl & Sinnott, 1991, Bergquist et al, 1992).

Leisure physical activity was not measured, and may have been another influencing source, as the mentioned bussing pattern. The opinions of the teachers were that organised physical activity was comparable in 2001 and 2002. There is a possibility that the experiences of walking at school may have induced more unorganised leisure activity among the children, like walking or bicycling for transport or for fun. As could be expected, the pupil's individual physical performance before the intervention influenced the effect of intervention, as the effect was highest among those needing it most. For every reduction of one unit of hamstrings flexibility, low back static endurance and heart rate before the intervention, the intervention effect was increased by 0.4, 0.2 and 0.3 units, respectively (Table 4).

Our data confirm theories from the Kraus-Weber-investigations, which postulated that a passive lifestyle of seldom walking and much time spent in cars, buses and in front of a television were the main causes of cross-cultural differences in physical performance in the 1950s between American and European children (Kraus & Hirschland, 1954). Our findings also support theories of a secular trend in European children of poorer physical performance as regards low back endurance (Kraus & Raab, 1961, Kraus & Hirschland, 1954, Sjolie & Mønness, 2006). Regular walking or bicycling was associated with increased low back endurance and hip mobility in a Norwegian cross-sectional study of 15-year old adolescents (Sjolie, 2000). School children who walked or bicycled 2-4 km one way to school had better cardio-vascular fitness than those walking or bicycling shorter distances in a large Norwegian study (Solstad, 1973).

Walking is postulated to enhance physical performance and prevent a variety of disorders (Morris & Hardman, 1997). Walking to school was inversely associated with low back pain in a Belgian prospective study (Szpalski, 2002). The increase of low back static endurance and hamstrings flexibility may reduce the risk for the development of low back pain, as these factors were inversely associated with low back pain in some studies (Sjolie & Ljunggren, 2001, Balague, 1999).

Studies of adults have also found health benefits from walking. Eight km walking on a golf course 2.5 times a week for 20 weeks increased low back static extension and cardio-vascular fitness, but not balance in a Finnish trial of sedentary 45-65 year-old males (Stensel, 1994). One hour walking or bicycling to work for ten weeks increased cardio-vascular fitness in inactive middle-aged Finnish men and women (Oja, 1991).

The need for more physical activity during childhood and adolescence seems to be urgent, as recent studies indicate a sedentary lifestyle in both youngsters and adolescents. A Scottish study following three year-old children two years ahead showed by using objective measures that median percentage time spent in moderate to vigorous physical activity of monitored waking hours was only two per cent at the age of three and four percent at the age of five (Reilly, 2004). A recent Norwegian study of 15-year old adolescents found that 45 % participated less than one hour per day in moderate physical activity (Klasson-Heggebo & Anderssen, 2003). The low levels of physical activity constitute risks also for e.g. development of obesity, cardio-vascular diseases, a high blood pressure, diabetes, type 2 and osteoporosis (Pate et. al, 1995). Various health promotion initiatives are planned in US to inspire people to increase light physical activity like some daily walking (http://www.Americaonthemove.org). Motivational strategies of walking are probably insufficient and should be accompanied by environmental planning in the communities of walking trails (Lumsdon & Mitchell, 1999).

Conclusion

The experimental design used here may be convenient in small-scale school investigations where an ordinary control group is not feasible or available. Our findings indicate good health effects of short, but frequent and persistent walking. The effects were largest among those with an initial poor physical performance. The model of daily walking for short periods during school time may thus be a relevant means to improve physical performance in children.

References

Altman DG. Practical statistics for medical research. Chapmann & Hall 1991. (reprint 1999)
Ahlberg A, Moussa M, Al-Nahdi M. On geographical variations in the normal range of joint motion. Clin Orth Rel Res 1988;234:229-31.
Alaranta H, Hurri H, Heliovaara M, Soukka A, Harju R. Non-dynamometric trunk performance tests: Reliability and normative data. Scand J Rehabil Med 1994;26:211-5.
Balague F, Troussier B, Salminen JJ. Non-specific low back pain in children and adolescents: risk factors. Eur Spine J 1999;8:429-438.
Beaglehole R, Bonita R. Public Health at the Crossroads. Cambridge University Press 1999.
Bergkvist M, Hedberg G Rahn M: Utværdering av test for bedømning av styrka, rørlighet och koordinasjon. Arbeta och Hælsa 1992-5 Arbetsmiljøinstitutet i Sverige, 1992. (Evaluation of tests for estimation of strength, mobility and coordination. Work and Health 1992-5. The Swedish National Institute for Working Life)
Bryk, A. S. & Weisberg H. I: Value-added analysis: A Dynamic approach to the estimation of treatment effects. Journal of Educational Statistics. 1976 Vol 1 No 2 Pp 127-155
Byl NN and Sinnott PL. Variations in balance and body sway in middle-aged adults: Subjects with healthy backs compared with subjects with low-back dysfunction. Spine 1991;16: 325-330.
Department of Transport. National travel survey 1992–4. HMSO, London 1995.
Ellingsen F. Kartlegging av styrke, bevegelighet, koordinasjon og utholdenhet. Skoleelever i Akershus 1968 og 1997. (A survey comparing strength, mobility, coordination and endurance. School children in 1968 and 1997. Norwegian language). The Norwegian University of Sport and Physical Education, Oslo 1999.
Everitt, B S(Author). Statistical Aspects of the Design and Analysis of Clinical Trials, Revised Edition. Imperial College Press 2004. Reference to chapter 5.
Gajdosik R and Lusin G. Reliability of an Active-Knee-Extension Test. Phys Ther 1983;63: 1085-1090.
Heck R. H. Examining the Impact of School Quality on School Outcomes and Improvement: A Value-Added Approach. Educational Administration Quarterly 2000; 36; 513
Hoaglund FT, Yau AC, Wong WL. Osteoarthritis of the hip and other joints in southern Chinese in Hong Kong. J B J Surg 1973;55A:545-57.
Johnsson B: Postural faults in school children. Thesis. Department of physiotherapy, Lund 1983.
http://www.Americaonthemove.org
Klasson-Heggebo L and Anderssen SA. Gender and age differences in relation to the recommendations of physical activity among Norwegian children and youth. Scand J Med Sci Sports 2003;13:1-6.
Kraus H and Hirschland R. Minimum muscular fitness tests in school children. Res Quart 1954:37:178-188.
Kraus H and Raab W. Hypokinetic disease. Thomas Publisher, Springfield. 1961.
Lockwood J. R. McCaffrey D. F. Hamilton L. S. Stecher B. Le V. Martinez J. F. The Sensitivity of Value-Added Teacher Effect Estimates to Different Mathematics Achievement Measures. Journal of Educational Measurement. 2007, Vol. 44, No. 1, pp. 47–67
Lumsdon L and Mitchell J. Walking, transport and health: Do we have the right prescription? Health Promotion International 1999;14: 271-279.
Morris JN and Hardman AE. Walking to health. Review. Sports Med 1997;23 306-32.

Oja P, Mänttäri A, Heinonen A, Kukkonen-Harjula K, Laukannen R, Paisanen M, Vuori I. Physiological effects of walking and cycling to work. Scand J Med Sci Sports 1991;1: 151-7.

Oksanen A and Salminen JJ. Tests of spinal mobility and muscle strength in the young: Reliability and normative values. Physiother Theor Pract 1996;12:151-60.

Pate RR, Pratt M, Blair SN, Haskell WL, Macera CA, Bouchard C, Buchner D, Ettinger W, Heath GW, King AC, et al. Physical activity and public health. A recommendation from the centers for disease control and prevention and the American college of sports medicine. Review. JAMA 1995;273(5):402-7.

Pettersen T. Måling av bevegelighet hos elever i grunnskolen. (Measurement of mobility in primary school children). Master's thesis. The Norwegian University of Sports and Physical Education, Oslo. 1984.

Reilly JJ, Jackson DM, Montgomery C, Kelly LA, Slater C, Grant S, Paton JY. Total energy expenditure and physical activity in young Scottish children: mixed longitudinal study. Lancet. 2004; 363:211-2.

Seltzer M. Choi K. Thum Y. M. Examining Relationships Between Where Students Start and how Rapidly they Progress: Using New Developments in Growth Modeling to Gain Insight into the Distribution of Achievement Within Schools. Educational Evaluation and Policy Analysis 2003, Vol. 25, No. 3, pp. 263–286

Sjolie AN and Ljunggren AE. The significance of high lumbar mobility and low lumbar strength for current and future low back pain in adolescents. Spine 2001;26(23):2629-36.

Sjolie AN. Pedestrian roads access, daily activities and performance in adolescents. Spine 2000;25:1965–1972.

Sjolie AN and Mønness E: Truncus endurance, hip and ankle mobility and aerobic fitness in 15-year old Norwegian adolescents in 1968 and 1997. Scandinavian Journal of Medicine & Science in Sports 2006. (OnlineEarly Articles). doi:10.1111/j.1600-0838.2006.00597.x

Solstad KJ. Skoleskyss og fysisk helse (School bus and physical health). (Dissertation. Norwegian text; English summary.) University of Tromso, Norway. 1973.

Stensel DJ, Brooke-Wavell K, Hardman AE, Jones PR, Norgan NG. The influence of a 1-year programme of brisk walking on endurance fitness and body composition in previously sedentary men aged 42-59 years. Eur J Appl Physiol Occup Physiol. 1994;68: 531-7.

SYSTAT 9.0. Copyright SPSS Inc. 1998.

Szpalski M, Gunzburg R, Balague F, Nordin M, Melot C. A 2-year longitudinal study on low pack pain in primary school children. Eur Spine J 2002;11:459-64.

Tomkinson GR, Leger LA, Olds TS, Cazorla G. Secular trends in the performance of children and adolescents (1980-2000): an analysis of 55 studies of the 20m shuttle run test in 11 countries. Sports Med. 2003;33:285-300.

Thum Y. M. Measuring Progress Toward a Goal: Estimating Teacher Productivity Using a Multivariate Multilevel Model for Value-Added Analysis. Sociological Methods Research 2003; 32; 153

Vibe N. Våre daglige reiser. Endringer i nordmenns reisevaner fra 1985 til 1992. (Our daily travels. Alterations in travel habits among Norwegians from 1982 to 1992. Norwegian text).Institute of communication, report 17, Oslo. 1993.

Watkins J. Step-tests of cardiorespiratory fitness suitable for mass testing. Br J Sports Med 1984;18: 84-9.

Westerstahl M, Barnekow-Bergkvist M, Hedberg G, Jansson E. G. Secular trends in body dimensions and physical fitness among adolescents in Sweden from 1974 to 1995. Scand J Med Sci Sports 2003;13:128-37.

Table 1. The measurement pattern. Y_{1j0} and Y_{2j1} are the measurements before and after the intervention on the same pupils, Y_{2j0} and Y_{3j1} are the measurements before and after the intervention on the same pupils, etc.

	School year 1	School year 2	School year 3	School year 4	School year 5	School year 6	School year 7
After intervention, *i=1*		Y_{2j1}	Y_{3j1}	Y_{4j1}	Y_{5j1}	Y_{6j1}	Y_{7j1}
Before intervention, *i=0*	Y_{1j0}	Y_{2j0}	Y_{3j0}	Y_{4j0}	Y_{5j0}	Y_{6j0}	Y_{7j0}

Table 2: Means of physical performance before and after the intervention by school year, and number of pupils

	Intervention	School year means 1	2	3	4	5	6	7	School mean	St. error of the mean
Balance,	before	7	10	10	15	16	21	31	16	1.51
Seconds	after		14	27	25	28	39	35	29	2.23
Hamstrings,	before	66	79	67	65	74	76	75	72	1.17
Degrees	after		74	84	73	80	78	80	78	1.12
Back endurance,	before	58	97	89	111	139	183	209	130	7.64
Seconds	after		97	160	138	148	188	191	155	8.10
Heart rate 0 min.	before	157	142	154	164	158	157	149	155	1.75
after step-test	after		154	144	143	150	143	143	146	1.66
Heart rate 1 min.	before	95	99	96	100	90	94	88	95	1.69
after step-test	after		83	81	81	85	82	80	82	1.38
Heart rate 2 min.	before	92	86	95	93	84	86	84	89	1.46
after step-test	after		73	75	74	80	77	75	76	1.27
No. of pupils	before	6+5	6+6	13+2	7+11	12+2	5+10	9+6	58+42	
Boys+ girls	after		6+5	6+6	13+2	7+11	12+2	5+10	49+36	

Table 3: Overall intervention effect mean of every $\Delta Y_{cj} = \left(Y_{cj1} - Y_{c-1j0}\right) - \left(\overline{Y_{c\bullet 0}} - \overline{Y_{c-1\bullet 0}}\right)$ on physical performance of the entire examined population: Means, Standard Deviations (SD), 95 % confidence interval (CI), and performance increase in per cent. "Hart rate *number*" means "Hart rate *number* minutes after step-test". The "% performance increase is $\overline{\Delta Y_{\bullet\bullet}} / \overline{Y_{\bullet\bullet 0}}$ The rightmost column shows performance increase if not corrected for age effect $\left(\overline{Y_{c\bullet 0}} - \overline{Y_{c-1\bullet 0}}\right)$. Generally this will overestimate the intervention effect.

				95% CI		% performance increase	% performance increase uncorrected of age effect
	Means	SD	P value	Lower	upper		
Δ Balance, seconds	11.11	2.09	0.00	6.95	15.27	69%	95 %
Δ Hamstrings, degrees	5.59	1.10	0.00	3.41	7.77	8%	10 %
Δ Back endurance, seconds	13.89	6.31	0.03	1.34	26.45	11%	31 %
Δ Heart rate 0	-8.99	1.49	0.00	-11.96	-6.03	-6%	-7 %
Δ Heart rate 1	-12.17	1.71	0.00	-15.57	-8.78	-13%	-16 %
Δ Heart rate 2	-12.03	1.54	0.00	-15.10	-8.97	-14%	-16 %
Δ Heart rate 0-1	-3.20	2.03	0.12	-7.23	0.83	5%	5 %
Δ Heart rate 2-1	-3.04	1.91	0.12	-6.84	0.76	5%	6 %

Table 4: Testing the intervention effect by school year, gender and initial performance. The "Estimate" is the regression coefficient of Y_{cj0} on ΔY_{cj}

Δ Balance	Estimate	df	F-ratio	p-value
Intervention		1	20.03	0.00
School year		5	1.25	0.29
Gender		1	0.29	0.59
Balance before	-0.35	1	3.08	0.08
Δ Hamstrings	**Estimate**	**df**	**F-ratio**	**p-value**
Intervention		1	42.43	0.00
School year		5	15.86	0.00
Gender		1	4.02	0.05
Hamstrings before	-0.41	1	30.74	0.00
Δ Back endurance	**estimate**	**df**	**F-ratio**	**p-value**
Intervention		1	11.81	0.00
School year		5	3.59	0.01
Gender		1	0.20	0.66
Endurance before	-0.23	1	5.87	0.02
Δ Heart rate 0	**estimate**	**df**	**F-ratio**	**p-value**
Intervention		1	13.68	0.00
School year		5	11.31	0.00
Gender		1	2.92	0.09
Heart rate 0 before	-0.34	1	19.25	0.00
Δ Heart rate 1	**estimate**	**df**	**F-ratio**	**p-value**
Intervention		1	44.84	0.00
School year		5	3.61	0.01
Gender		1	5.92	0.02
Heart rate 1 before	-0.64	1	72.92	0.00
Δ Heart rate 2	**estimate**	**df**	**F-ratio**	**p-value**
Intervention		1	34.78	0.00
School year		5	5.19	0.00
Gender		1	8.71	0.00
Heart rate 2 before	-0.61	1	59.00	0.00
Δ(Heart rate 1-heart rate 0)	**estimate**	**df**	**F-ratio**	**p-value**
Intervention		1	76.51	0.00
School year		5	12.34	0.00
Gender		1	1.00	0.32
(Heart rate1 –Heart rate0) before	-0.85	1	67.44	0.00
Δ(Heart rate2-Heart rate0)	**Estimate**	**df**	**F-ratio**	**p-value**
Intervention		1	51.75	0.00
School year		5	8.05	0.00
Gender		1	0.34	0.56
(Heart rate2 –Heart rate0) before	-0.73	1	45.41	0.00

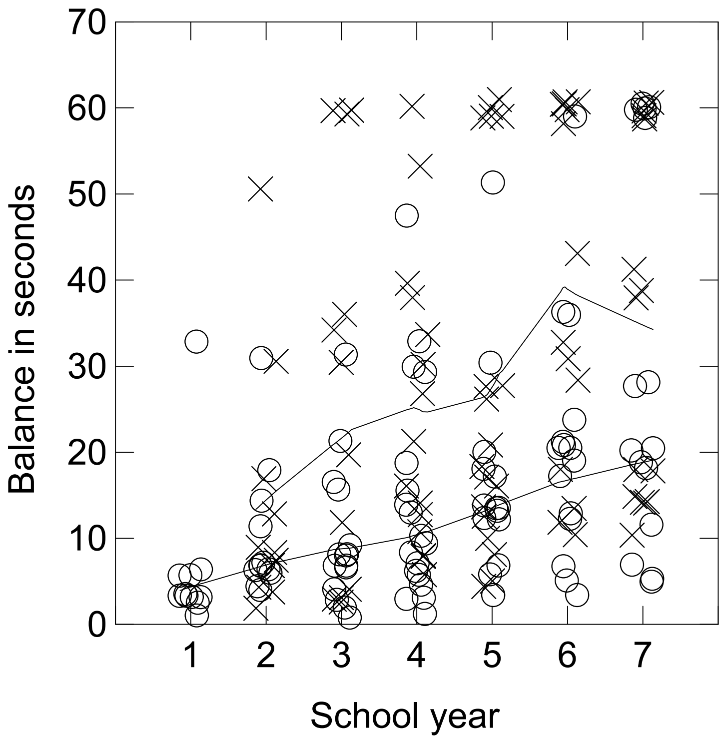

Figure 1 depicts the individual balance performance in seconds (Vertical axis). The measurements were taken up to 60 seconds. The horizontal axis is school year. The circles are before intervention measurements, the crosses are after intervention.. LOWESS regression lines are introduced to pinpoint the ageing effect. The gap between the lines indicates an intervention effect.

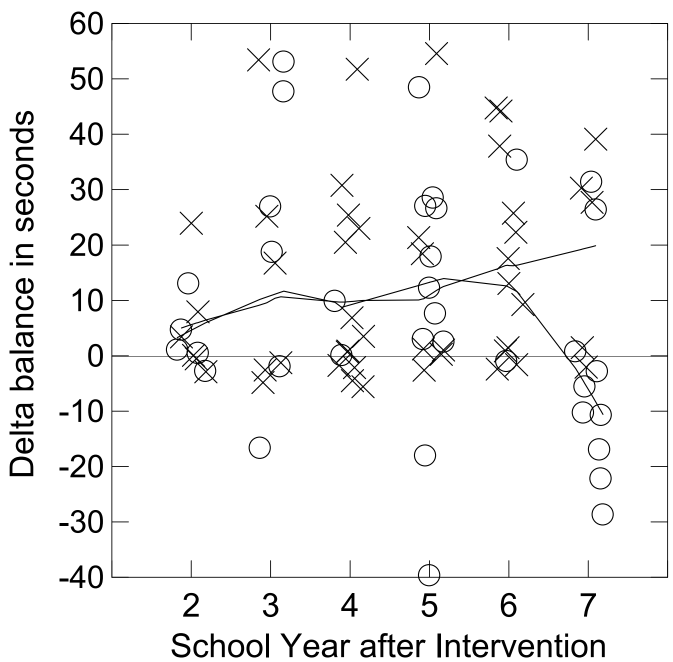

Figure 2 depicts ΔY_{cj}, the individual age corrected balance improvement (Vertical axis) . Horizontal axis is school year after intervention. The circles are girls, the crosses are boys. LOWESS regression lines are introduced separately for boys and girls. The size of the $\left(\overline{Y_{c\bullet 0}} - \overline{Y_{c-1\bullet 0}}\right)$ correction is by year: 3.5 -0.2, 4.6, 1.7, 4.7 and 9.8 for year 7.