

ORIGINAL ARTICLE

A useful evaluation design, and effects of the Olweus Bullying Prevention Program

DAN OLWEUS

Research Centre for Health Promotion (HEMIL), University of Bergen, Norway

Abstract

The article presents the logic and other characteristics of an “extended selection cohorts” quasi-experimental design. Possible threats to the validity of conclusions based on this kind of design are discussed. It is concluded that chances are good that conclusions about the effects or non-effects of school-based intervention programs will be roughly correct in most cases. The design may be particularly useful in studies where it is not possible or desirable to use a random selection of “control schools” and it should be of value to both practitioners and researchers. The design is illustrated with a study in which three consecutive cohorts of students (n approximately 21 000) were administered the Bully/Victim Questionnaire before and after some 8 months of intervention with the Olweus Bullying Prevention Program (OBPP). Results indicated quite substantial reductions (by 32–49%) in bully/victim problems. The “time-series” nature of the data showed convincingly that a “history interpretation” of the findings (Cook & Campbell, *Quasi-experimentation*. Chicago: Rand McNally, 1979) is very unlikely. The data in this project were obtained in the context of a government-funded new national initiative against bullying in Norway. The characteristics of this initiative and the model used in implementing the program in more than 450 schools were briefly described.

Keywords: *Selection cohorts design, evaluation, prevention, bully, victim*

Introduction

A researcher or practitioner who is interested in evaluating the effects of an intervention program, for example against bully/victim problems in school, is very often faced with a situation where it is not possible or desirable to use a traditional experimental design. This means that the observational units such as students or classes/schools are not randomly assigned to the various treatment conditions (e.g. intervention versus no intervention/control, or various degrees of intervention versus no intervention/control). In such situations, the investigator usually must turn to what is called a quasi-experimental design. How can the investigator then evaluate the effects of an intervention in a reasonably rigorous way?

Correspondence: Dan Olweus, Research Centre for Health Promotion (HEMIL), University of Bergen, Christies Gate 13, N-5015 Bergen, Norway. E-mail: Olweus@psyhp.uib.no and bvpolweus@c2i.net.

There is a large literature on various quasi-experimental designs (see, e.g. Cook & Campbell, 1979; Shadish, Cook, & Campbell, 2002, and a number of standard textbooks in design and statistics) the strength and weaknesses of which will not be discussed in the present context. However, here I will focus brief attention on one particular design which I have found particularly useful and which is relatively easy to use, also for investigators who are not primarily researchers. The general structure of this design, sometimes called a *selection cohorts design* is described in Cook and Campbell's classical book (1979) and some textbook treatments (e.g. Judd & Kenny, 1981, under the name of "age cohort design"). However, one does not see many examples of it in the literature, in particular not the "extended" version of the design (below) which I recommend and have used in several intervention studies. Important aspects of this variant of the design are that it consists of several adjacent or contiguous cohorts and that there is a 1-year (or possibly 2-year) interval between measurement occasions.

A concrete illustration

I will start by giving a brief description of this extended version as it was used in the First Bergen Project against Bullying (e.g. Olweus, 1991, 1993, 1994a). Since this project was part of a nationwide campaign against bullying, it was not possible to set up a strictly experimental study with random allocation of schools or classes to treatment and control/comparison conditions.

Evaluation of the effects of the intervention program was based on data from approximately 2500 students who were followed over a period of 2.5 years. The students originally (at Time 1, below) belonged to 112 grade 4–7 classes (corresponding to grades 5–8 in the new grade system) in 42 primary and junior high schools in Bergen. Each of the four grade/age cohorts (with modal ages of 11, 12, 13, and 14 years, respectively, at Time 1) consisted of 600–700 subjects with a roughly equal distribution of boys and girls. In the present context, the students belonged to a "cohort" in the sense that they were joined together in distinct classes within a particular grade level and were approximately the same age. The first time of data collection (Time 1) was in May/June 1983, approximately 4 months before the intervention program was introduced in October. New measurements were taken in May 1984 (Time 2) and May 1985 (Time 3). The intervention program was (more or less) in place for the whole 20-month period from October 1983, until May/June 1985. The basic structure of the design is shown in Figure 1 (for ease of exposition and understanding, the figure uses fictitious and idealized data which to some extent reflect the general trends of the empirical findings for "being bullied"; however, with regard to "involvement in antisocial behavior", for example, the expected developmental curves would go upwards).

For three of the cohorts (C5, C6, and C7), data collected at Time 1 were used as a baseline with which data for age-equivalent cohorts at Time 2 could be compared. The latter groups had then been exposed to the intervention program for about 8 months. To exemplify, the data for the grade 5 cohort at Time 1 (modal age 12 years) were compared with the Time 2 data for the grade 4 cohort which at that time had reached approximately the same age as the baseline group. The same kind of comparisons were made between the grade 6 cohort at Time 1 and the grade 5 cohort at Time 2, and between the grade 7 cohort at Time 1 and the grade 6 cohort at Time 2.

Comparison of data collected at Time 1 and Time 3 permitted an assessment of the persistence or possible decline or enhancement of the effects over a longer time span. For

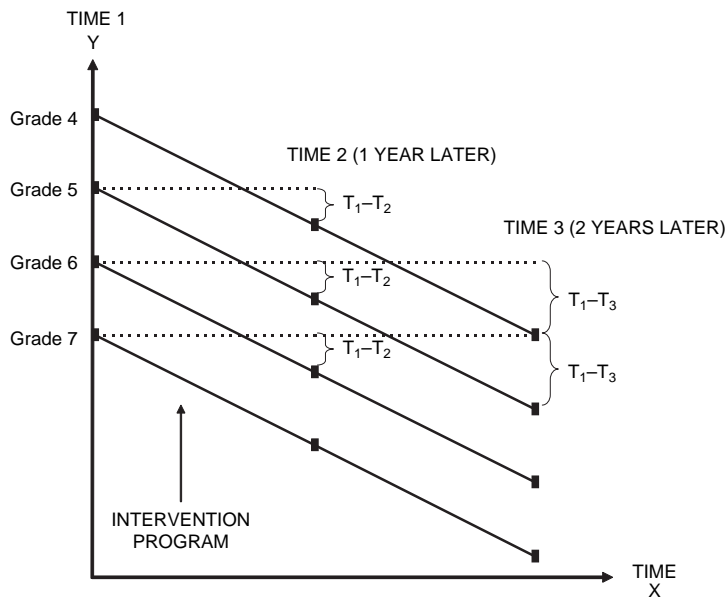


Figure 1. Design for evaluation of effects of the intervention program. Fictitious data (which to some extent reflect the general trend of the empirical findings for “being bullied”) are shown.

these comparisons data for only two of the cohorts could be used as a baseline, those of the grade 6 and grade 7 cohorts, which were contrasted with data collected at Time 3 on the grade 4 and grade 5 cohorts, respectively. The latter groups had been exposed to the intervention program during approximately 20 months at that time.

Additional characteristics of the design

In any study designed to establish or make probable the effects of some factor such as an intervention program, it is mandatory that the investigator examines and ideally is able to rule out most or all alternative explanations of the findings in terms of possible confounding, “irrelevant” factors. This is true whether the study is experimental or quasi-experimental, although it is obvious that certain alternative interpretations can be more easily eliminated if the units of sampling have been allocated to the various conditions by a random procedure. Accordingly, it is very important to be aware of common possible “threats to the internal validity” (Cook & Campbell, 1979) of any design, and to examine to what extent and in what ways such threats or sources of confounding can possibly be eliminated or counteracted. Here I will only give a brief discussion of these issues in relation to the present design. For more detailed and somewhat more technical discussions of certain aspects, see Olweus & Alsaker (1991) and Olweus (1991).

A key aspect of this design is that the relevant groups or cohorts compared are the same age. This is necessary in order to take care of or rule out explanations of the results in terms of *differences in age* or “*maturation*”. It is well documented that there occur changes in bully/victim problems as a function of age (e.g. Olweus, 1993; Smith, Madsen, & Moody, 1999; Solberg & Olweus, 2003). Accordingly, such *developmental changes* must be “controlled”, and this is done by comparison of age-equivalent groups at the various time points.

By controlling for age in this way, the time of the year when the outcome or dependent variable(s) is being measured is also “held constant”. In some areas such as bully/victim problems, this may be important in order to control for possible seasonal variations (due to the amount and nature of outdoor activities, typical interaction patterns, etc.).

A major problem in many quasi-experimental studies relates to the fact that *the intervention group(s) and control or comparison groups differ in known or partly unknown ways in important aspects* before the intervention is introduced. And, as has been repeatedly pointed out in the statistical literature (e.g. Miller & Chapman, 2001; Porter & Raudenbush, 1987), the common strategy of using analysis of covariance, ANCOVA, to “control for” initial differences among pre-existing groups is often an inappropriate or risky enterprise. Accordingly, it is a great advantage if the investigator can get hold of *naturally occurring groups that are reasonably similar or equivalent* with regard to the outcome variable (and dimensions related to the outcome variable), *before* the intervention is administered to one or more of the groups. When the groups to be compared belong to the same schools (for example, the grade 5 cohort at Time 1, with no intervention, compared with the grade 4 cohort at Time 2, with 8 months of intervention, recruited from the same schools), there are often good grounds for assuming that one cohort differs in only minor ways from its contiguous cohort(s). Usually, the majority of the members in the various grade cohorts have been recruited from the same relatively stable populations and have also been students in the same schools for several years. In some cases, in particular in more research-oriented studies, it may be possible to check the similarity of the groups compared with regard to presumably important dimensions.

In spite of the fact that cohorts selected from the same schools can often be assumed to be reasonably equivalent in important respects, it is possible that some kind of *selection bias* can occur. Such bias could be the result of inadvertent changes in the recruitment of students to the various cohorts so that the cohorts in fact represent populations with partly different compositional characteristics. If present, such bias might complicate interpretation of the time-lagged comparisons. However, here it is important to emphasize that the extended selection cohorts design with adjacent cohorts provides *partial protection against such selection bias*. This is due to the fact that several of the cohorts serve as a baseline group in one set of comparisons and as an intervention group in another. This is the case, for example, with the grade 5 cohort at Time 1, the data for which are used as a baseline in comparison with the grade 4 cohort data collected at Time 2 (after 8 months of intervention; see Figure 3). At the same time, the grade 5 cohort data obtained at Time 2 serve to evaluate the possible effects of 8 months of intervention when they are compared with the data for grade 6 cohort at Time 1. The same situation applies to grade 6 cohort in comparisons with grades 5 and 7 cohorts, respectively.

The considerable advantage of this aspect of the design is that a possible bias in the composition of the cohorts would operate in opposite directions in the two sets of comparisons, thus making it difficult to obtain apparent “intervention effects” across cohorts as a consequence of such selection bias. This feature of the design also provides protection against faulty conclusions in case the baseline data for one or both of these cohorts were unusually high or low simply as a function of chance. The protection against selection bias is partial in the sense that the both the youngest and the oldest cohort, in the present illustration the grade 4 cohort and the grade 7 cohort, only serve as an intervention group (grade 4 cohort at Time 2) or a baseline group (grade 7 cohort at Time 1) with regard to the Time 1 – Time 2 comparisons.

To safeguard against erroneous conclusions due to *possible selective attrition* (for example, more extreme or deviant individuals may be more likely to drop out in longitudinal studies), analyses can be restricted to students for whom there are valid data at both time points in a particular comparison (both for the baseline and the intervention groups).

In addition, it should be mentioned that selection of groups/subjects in this design is typically not based on some kind of “extreme score” criterion. Accordingly, the problem with “regression toward the mean”, which looms large in many evaluation studies, is not an issue here.

Possible effects of repeated measurement and “history”

However, there are two additional conceivable sources of confounding that must also be considered. One relates to possible “*testing*” or *repeated measurement effects*. As evident from Figure 1, the scores for the baseline (Time 1) data usually represent a first-time measurement, whereas the Time 2 data come from a second wave of measurement. Although it may not appear very likely that a second measurement, separated by a whole year from the first measurement occasion, would result in some kind of systematic change in the students’ responding, it may, by way of precaution, nevertheless be valuable to examine if such changes have occurred and in what directions they might go. If such (nontrivial) effects were found, this might complicate interpretation of the results. One way of examining such effects would be to let half of the students/classes in the youngest cohort skip the first measurement occasion, and then compare the two halves of this cohort at Time 2. In the First Bergen Project, the possibility of repeated measurement effects was investigated in a slightly different way (see Olweus, 1991, p. 442) and such effects were found to be small and non-systematic.

With a selection cohorts design, one has also to be aware of the possibility that registered changes in the outcome variable are a consequence of some “*irrelevant*” factor concomitant to the intervention program, implying that the results may be given a “*history*” interpretation (see Cook & Campbell, 1979). It might be, for example, that the intervention groups were exposed, in addition to the intervention program, to some kind of changes in the educational, administrative, or other school routines which may have affected their behavior and responding at Time 2. Accordingly, it may be important for the investigator to examine if such parallel changes have occurred during the intervention period(s) and if so, whether they can be meaningfully linked to systematic changes in the outcome variable(s). A similar argument can be made with regard to “*general time trends*” in the outcome variable or related dimensions, that is, historical societal changes (often due to unknown causes) which happened to coincide with the intervention (see Olweus, 1994b, pp. 120–121). Although such “*history explanations*” frequently appear fairly unlikely, particularly in consideration of the relative abruptness of the changes often observed, the investigator can obtain additional help in ruling out (or possibly incorporating) such interpretations if he or she can also include in the design some equivalent units (schools/classes) without any intervention at all, that is, some “*control*” units.

Although the extended selection cohorts design has many attractive features, there are also some limitations that deserve mention. First, it may be noted (as indicated above) that some of the collected data can not be used in the evaluation of the program effects. This is true of the grade 4 cohort data at Time 1 and the grade 7 cohort data at Time 2, for example, with regard to the Time 1 – Time 2 comparisons. Also, although the design is longitudinal, this aspect can not be taken into account in the statistical analyses. This means

that the advantage of having repeated measurements on the same subjects is not translated into a reduced error term. Accordingly, the design is likely to have less statistical power or precision than if a repeated-measures design had been used. At the same time, these two concerns may not be very important in the context of a selection cohorts design where large amounts of data can often be collected without great effort.

It should be mentioned that all of these possible alternative explanations (in addition to potential under- and over-reporting) of the systematic reductions in bully/victim problems and related behavior patterns (below) were carefully examined in the First Bergen Project against Bullying and generally found to be deficient in explaining the results obtained (Olweus, 1991, 1993; Olweus & Alsaker, 1991). In addition, a clear “dosage–response” relationship ($r = 0.51$, $n = 80$) was established in preliminary analyses at the class level (which is the natural unit of analysis in this case): those teachers/classes that had larger reductions in bully/victim problems had implemented three presumably essential components of the intervention program (including establishment of class rules against bullying and use of regular class meetings) to a greater extent than those with smaller changes. This finding provides corroborating evidence for the hypothesis that the changes observed were a consequence of the intervention program and not of some other “irrelevant” factor.

Some of the arguments presented above may appear somewhat subtle and technical. They are, however, quite important to consider in a research study, whether experimental or quasi-experimental, aiming to document the possible effects of an intervention program. In addition, for an adequate statistical treatment, the hierarchical or “nested” nature of the data must be taken into account (see Olweus, 1991; Olweus & Alsaker, 1991). However, the extended selection cohorts design has a number of attractive features and built-in safeguards which should facilitate interpretation of the results. In addition, practitioners such as the school leadership or a school board, can probably take most of the validity concerns discussed fairly lightly, provided that the intervention situation is reasonably “clean”. By this I mean that preferably no other intervention programs or similar activities or events *are introduced* in the participating schools in the same time period as the program at issue is being evaluated. (In preparation of possible implementation of the Olweus Bullying Prevention Program at a particular school, we very strongly advise the school leadership not to start implementation of some other program at the same time, both in consideration of the necessary time and energy resources, possible negative interactions among programs, and likely ambiguities with regard to interpretation of possible “intervention effects”. In case the school concerned already has in place a program which in one way or another is in conflict with the principles and general approach of the Olweus Program, the school leadership is strongly recommended to postpone implementation of the Olweus Program to a later point in time.)

Summing up, with use of an extended selection cohorts design of the type described above, chances are quite good that conclusions about the effects or non-effects of an intervention program *will be roughly correct in most cases*. In addition, the design is easy to use; it is actually a very natural step in the monitoring of what goes on in schools involved in anti-bullying work. Overall, the design must be considered *quite a useful design for practitioners and researchers alike*, and in the author’s view, it is clearly underused. However, the many positive aspects of the design can not, of course, exempt us from the responsibility of using other available data and our heads in making a balanced evaluation of the results obtained.

Some evaluations of the effects of the Olweus Bullying Prevention Program (OBPP)¹

As described above, an extended selection cohorts design was used to evaluate the effects of the intervention program in the First Bergen Project against Bullying, running from 1983 to 1985. Variants of the same design, with some modifications, have also been used in two more recent evaluation projects: the New Bergen Project against Bullying from 1997 to 1998 (1999a) and the New National Initiative against Bullying with evaluation data from 2001 to 2003 (so far).

The New Bergen Project against Bullying comprised some 3200 students in grades 5–7 and 9 belonging to 14 intervention and 16 “comparison” schools (Olweus, 1999a). The fact that the second group of schools was called comparison schools did not in any way imply that they were not involved in some kind of intervention work against bullying (which is actually expected from all schools in present-day Norway). However, they were not part of the intervention project involving the Olweus Bullying Prevention Program (OBPP).

The New National Initiative against Bullying (to be described in somewhat more detail below) comprised more than 100 schools with approximately 21 000 students in grades 4–7. These schools applied for participation in the OBPP at three different time points, the autumn of 2001, the spring of 2002 and the autumn of 2002 when they also took the Olweus Bully/Victim Questionnaire (Olweus, 1986, 1996) for a baseline assessment. The second measurement with the same instrument occurred 1 year later, when the schools had worked with the program for approximately 8 months. This project is particularly interesting from the point of view that it can shed light on possibly confounding effects of general time trends or media attention.

Here I will only give a relatively brief summary of the results from these three projects, without statistical details. It should be noted, however, that in all statistical analyses, the hierarchical or nested nature of the data (with students nested within classrooms nested within schools) was taken into account (e.g. Olweus, 1991; Olweus & Alsaker, 1991; Raudenbush & Bryk, 2002). All main conclusions are based on results that are statistically significant or highly significant. For the two recent studies, I will restrict the reporting to data from the elementary grades (5–7 and 4–7, respectively) where important components of the program were more fully implemented than in higher grades.

Results from the two Bergen projects

The main results from the First Bergen Project can be summarized as follows (e.g. Olweus, 1991, 1993; Olweus & Alsaker, 1991):

- There were marked (and statistically highly significant) reductions—by 50% or more—in self-reported bully/victim problems for the periods studied, with 8 and 20 months of intervention, respectively. By and large, the results applied to both boys and girls and to students from all grades studied. Similar results were obtained for a kind of aggregated peer rating variables and teacher ratings. However, the effects were somewhat weaker for the teacher ratings.
- There were also clear reductions in general antisocial behavior such as vandalism, fighting with the police, pilfering, drunkenness, and truancy.
- In addition, we could register marked improvement as regards various aspects of the “social climate” of the class: improved order and discipline, more positive social

relationships, and a more positive attitude to schoolwork and the school. At the same time, there was an increase in student satisfaction with school life.

In the New Bergen Project against Bullying (1997–1998), we could again register clear improvements with regard to bully/victim problems in the intervention schools but the effects were somewhat weaker than in the first project, with averages varying between 21 and 38%. It should be noted, however, that the intervention program had been in place for only 6 months or less when the second measurement was made (at Time 2). In addition, this particular (1997/98) year was a very turbulent one for the teachers with the introduction of a new National Curriculum which made heavy demands on their time and emotional resources. For the comparison schools, there were very small or no changes in being bullied and actually an increase in the level of bullying other students by about 35%. Without having analyzed the questionnaire information obtained from the teachers in the comparison (and intervention) schools, we are not prepared to give a detailed explanation of this result. However, it is certainly consistent with findings from a number of studies which have found negative effects of interventions intended to counteract delinquent and antisocial behavior (e.g. Dishion, McCord, & Poulin, 1999; Gottfredson, 1987; Lipsey, 1992).

Results similar to those of the two Bergen projects, although somewhat weaker, have been obtained in partial replications in the UK, the USA, and Germany (Olweus & Limber, 1999; Smith & Sharp, 1994).

Stevens, de Boudeaudhuij, and Van Oost (2000) have raised the question of whether the positive results of the First Bergen Project against Bullying could possibly be at least partly a so-called Hawthorne effect, that is, a consequence of general attention from the media and the general public rather than an effect of the intervention program itself. As detailed in a publication written in Swedish (Olweus, 2002), I argue that this hypothesis, for a number of reasons, is highly unlikely to be true. For lack of space, I will limit myself to listing some of the arguments against the hypothesis. They focus on (a) the timing of the media attention to the Project, (b) the interpretation of allegedly contradictory results from a study by Roland (1989; Olweus, 1999b), (c) the nature of a possible Hawthorne effect, (d) the breadth of the program effects, and (e) the documentation of a “dosage–response” relationship mentioned above. In addition, the results from the project to be presented next clearly contradict such an interpretation.

Results from the New National Initiative Project

The main results of the New National Initiative Project can be seen in Figures 2 and 3. To better understand the meaning of the curves, let us focus on Figure 2. The upper curve portrays the baseline (before intervention) percentages of bullied students in grades 4–7 from five different cohorts of schools who participated in the OBPP for 18 months. To be classified as being bullied, the student had to respond to the global question in the Olweus Bully/Victim Questionnaire (Olweus, 1986, 1996) that he or she had been bullied “2 or 3 times a month” or more in the past couple of months. Psychometric analyses have shown that this is a useful and reliable way of dichotomizing the global variable (Solberg & Olweus, 2003). In other analyses, the full variability of the five-point global variable and various scales or indices have been used, and with similar results.

At the time of writing this article, we mainly focus on the first three of these cohorts for which complete follow-up data 1 year later were available. The percentages of students who

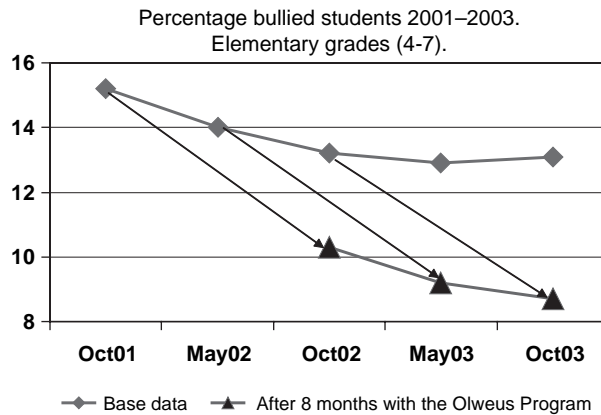


Figure 2. Percentages of bullied students, 2001–2003, before (upper curve) and after (lower curve) intervention with the Olweus Bullying Prevention Program.

reported being bullied 1 year later when the schools had used the OBPP for approximately 8 months are shown in the lower curve. Data from a particular cohort of schools are connected with an arrow. To illustrate, the percentage of bullied students in the first cohort of schools ($n = 8388$) was 15.2 while at follow-up 1 year later, this percentage had been reduced to 10.3—a relative reduction of 32%. The relative reductions for the two successive cohorts of schools were very similar, both amounting to 34% ($n = 4083$ and $n = 8238$). Absolute reductions for these three cohorts amounted to 4.9, 4.8, and 4.5 percentage points, respectively.

In Figure 3, the variable portrayed is bullying other students “2 or 3 times a month” or more in the past couple of months. The general pattern of results is very similar to what was reported for being bullied in Figure 2, but at a lower level, as expected. The relative reductions for the first three cohorts of schools (the same as in Figure 2) were 37%, 48% and 49%, respectively. The absolute reductions amounted to 2.1, 2.8, and 2.5 percentage points.

Generally, the results for both outcome variables were very similar across the three cohorts. There were very substantial relative reductions in the levels of bully/victim

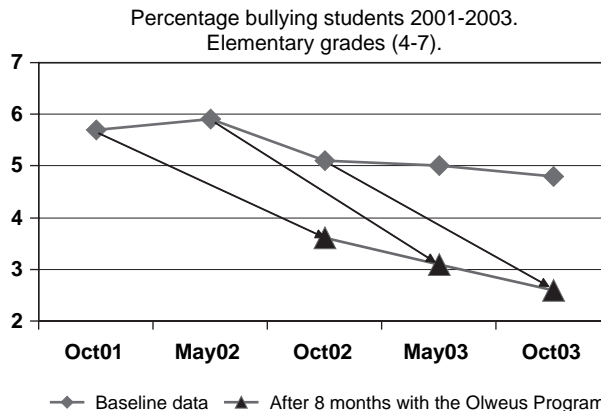


Figure 3. Percentages of bullying students, 2001–2003, before (upper curve) and after (lower curve) intervention with the Olweus Bullying Prevention Program.

problems, ranging between 32% and 34% for being bullied, and between 37% and 49% for bullying other students. In terms of absolute reductions, the figures varied between 4.5 and 4.9 percentage points for being bullied, and between 2.1 and 2.8 percentage points for bullying other students. The fact that the relative reductions were greater for bullying others in spite of the fact that an opposite pattern was found for absolute reductions is, of course, due to the fact that the baseline values from which the relative reductions were derived were smaller in the case of bullying others.

Figures 2 and 3 show percentages for boy and girl samples combined across grades 4–7. Basically similar results were obtained when the data were analyzed separately for the two genders, the four grades, and when a stricter criterion—“about once a week” or more often—was used in classifying students as being bullied or bullying other students. Marked improvements could thus be registered also for students who had been involved in more serious bully/victim problems. (It should be noted that having been bullied/bullied other students “2 or 3 times a month” by no means represents non-serious or trivial problems, as shown in Solberg & Olweus, 2003.)

When the percentages of change reported above were calculated, we did not just follow the same subjects over time and calculate the degree of change for each participant from baseline to follow-up. As explained in the first part of the paper, the key comparisons were being made between age-equivalent groups, that is, the data for grade 6 students at follow-up (after 8 months of intervention), for example, were compared with the baseline data (before intervention) for the grade 6 students in the same schools. The same procedure was followed for the other grades.

In more detailed analyses of the results from one of the cohorts (so far), we could register a number of changes in other areas or dimensions which also strongly suggested that the positive results were a consequence of the intervention. As an illustration, the students reported more active intervention in bullying situations from both teachers and peers at follow-up as compared with baseline. Also, at follow-up there were clearly more students who responded that the homeroom/main classroom teacher had done “much” or “a good deal” to counter bullying in the classroom in the past few months.

Is a “history interpretation” reasonable?

In explaining the logic of the extended selection cohorts design, a “history interpretation” was mentioned as one possible threat to validity, implying that the researcher must try to rule out or minimize the possibility that general time trends or some “irrelevant” factor concomitant to the intervention could account for the results. The fact that several consecutive cohorts of schools were measured before intervention in the present project (upper curves in Figures 2 and 3) can shed a special light on the reasonableness of such alternative explanations.

We see then that there is a slight decline in the levels of being bullied and possibly bullying others for the first three time points. This slight decline could possibly be an indication of a general time trend or an effect of general media/public attention to bully/victim problems in Norwegian schools during the particular time period. However, in particular since most of the media attention to these problems started in the early autumn of 2002, an equally likely interpretation is that the schools that first applied for participation in the OBPP had slightly higher levels of problems than schools that came into the program at a somewhat later point in time. In any case, the possible decline across the first three cohorts was quite small and was also somewhat reflected in the outcome curves after 8 months of intervention (lower

curves in Figures 2 and 3). These results derived from different but reasonably comparable cohorts of schools indicate considerable stability over time in the average amount of bully/victim problems. We have also found similar levels of stability when individual or groups of schools have been followed over time (at 1-year intervals) without intervention.

All of these results suggest that, without systematic and effective intervention, the levels of bully/victim problems characterizing consecutive, largely comparable cohorts of schools at different time points or a cohort of schools followed over time, will be relatively stable at least for a period of a couple of years. This also implies that a “history interpretation” in terms of general time trends or special media attention cannot reasonably be invoked as an explanation of the positive changes in our intervention schools in the New National Initiative Project.

It should be emphasized in this context that, although the schools seeking participation in the OBPP were “self-selected”, the levels of problems characterizing the participating schools (before intervention) were clearly within the ranges typically found for nationally representative samples (e.g. Tikkanen & Junge, 2004) for approximately the same periods. It may be useful to know that the schools participating in the OBPP were thus not in any way “atypical” with regard to problem levels.

For approximately half of the schools from the first cohort we also obtained follow-up data (32 schools with about 4000 students) 2 years after the first measurement occasion and approximately 6 months after the implementation phase of the OBPP had ended. These data showed that the reductions gained after 1 year were maintained or even slightly increased after 2 years. These schools were roughly similar to the total cohort of schools in terms of problem levels at Time 1, and degree of reduction of problems between Time 1 and Time 2. The results indicate that the Time 1 – Time 2 reductions were not a temporary and short-lived phenomenon contingent on constant, “full resource” participation in the OBPP. Although the follow-up period (so far) was relatively limited, the findings suggest that the schools may have actually changed their culture, readiness and competence to deal with and prevent bully/victim problems in a more long-term way.

While systematic use of the OBPP with students in grades 4–7 has consistently produced very positive results, which seem to be relatively unique in an international perspective (see Smith, Pepler, & Rigby, 2004), it should also be mentioned that the effects have been more variable with students from lower secondary/junior high school grades. In about half of our evaluation projects, results with students in that age range have been less successful than with younger students. We think we know several of the reasons for these results which, however, will not be discussed in this context for lack of space. We are presently engaged in efforts to adapt the program, or rather its implementation, in order to achieve more consistently positive results also for these age groups.

A new national initiative against bullying in Norway

Against this background, it is natural to give a brief description of the development and organization of the project that the results presented in the previous sections were derived from.

In late 2000, the Department of Education and Research (UFD) and the Department of Children and Family Affairs (BFD) decided that the OBPP was to be offered on a large-scale basis to Norwegian elementary and lower secondary schools over a period of years. In building up the organization and infrastructure for this national initiative, two overriding principles guided our work: (1) to try to ensure that the program was implemented

according to the intentions of the program designer, that is, with reasonable fidelity (quality control); (2) to try to get the program implemented in a reasonable number of schools/communities in a relatively limited period of time, say 5 or 6 years.

To accommodate both of these principles at the same time, we use a four-level strategy of dissemination, a kind of “train-the-trainer” model. The Olweus Group against Bullying and Antisocial Behavior at the HEMIL Centre at the University of Bergen, trains and supervises specially selected *instructor candidates* who each train and supervise “key persons” from a number of schools (ideally about five schools per instructor candidate). These key persons are then responsible for leading recurrent “staff discussion groups” at each participating school. The basic structure of the model is shown in Figure 4.

The training of the instructor candidates consists of 10–11 whole-day assemblies distributed over a period of some 16 months. In between the whole-day meetings the instructor candidates receive ongoing consultation via telephone or email with members of my group. After having successfully completed the training period, they will be assigned status as certified Olweus trainers. (In implementing this “train-the-trainer” model in the USA, some modifications have been made to accommodate cultural differences and practical constraints. In particular, the number of whole-day assemblies has been reduced to four or five, and the Bullying Prevention Coordinating Committees at the individual schools have been accorded somewhat greater responsibility than in Norway.)

An important task for the trainer candidates is to hold a 2-day training with special *key persons* from each participating school (or in the USA, with members of the coordinating committee; see Olweus & Limber, 1999). The trainer candidates are also involved in the administration of the Bully/Victim Questionnaire (Olweus, 1996; Solberg & Olweus, 2003) and in interpreting and communicating the results to the individual school. The Questionnaire survey is an important vehicle for creating awareness and involvement among staff, students, and parents. In addition, the key persons receive continuing supervision and assistance from their trainer candidates.

Establishment of *staff discussion groups* at each participating school is an important tool for effective dissemination and implementation of the program. These groups with up to 15 participants meet regularly for approximately 90 minutes every other week under the

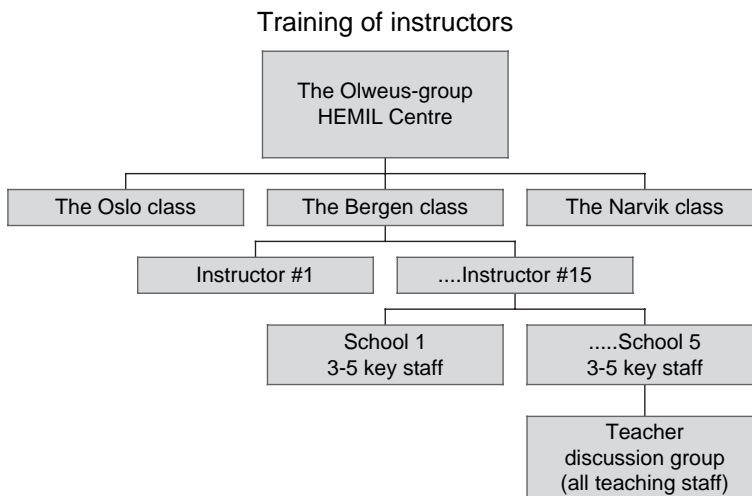


Figure 4. Overview of the “train-the-trainer” model used in the new national Norwegian anti-bullying initiative.

leadership of the specially trained key persons. The meetings are typically organized around important components or themes of the program as described in *Olweus' Core Program against Bullying: A Teacher Handbook* (Olweus, 2001) and the book *Bullying at School: What We Know and What We Can Do* (Olweus, 1993). The main goals of these meetings are the following:

- To provide more detailed and comprehensive knowledge of the intervention program and its various components.
- To provide the participants with the possibility of testing, through role playing and in other ways, ideas and practical solutions to various problem situations in a secure environment.
- To stimulate fast(er) implementation of the various components of the program.
- To share experiences and viewpoints with others in similar situations and to learn from others' positive and negative experiences.
- To create and maintain motivation and commitment.
- To stimulate cooperation and coordination of program components and activities (to develop and maintain a whole-school policy).

Although staff discussion groups may be perceived by some in the school society as rather time-/resource-consuming, the informal feedback we have received so far certainly suggests that these meetings are seen as very valuable by most participants. In many ways, these meetings around the program actually serve to stimulate organizational development of the school. A distinct advantage here is that the major goal of this form of school development is directed towards the students: to create a safe and positive learning environment.

Up to now, some 125 instructor candidates have finished or are in training, and more than 450 schools from all over Norway participate in the program. We perceive all of this as a breakthrough for the systematic, long-term, and research-based work against bully/victim problems in school and hope to see similar developments in other countries.

Acknowledgements

The research program reported on in this article was supported by grants from the Ministry of Children and Family Affairs (BFD), the Ministry of Research and Education (UFD), and in earlier phases from the Research Council of Norway (NFR, NAVF) and the Johann Jacobs Foundation, Switzerland, which is gratefully acknowledged.

Note

- 1 The intervention "package" consists of the book *Bullying at School: What We Know and What We Can Do* (Olweus, 1993; this book is sold in bookstores or through direct order from the publisher: Blackwell, 108 Cowley Road, Oxford OX4 1JF, UK, or its North American division: Blackwell, 238 Main Street, Cambridge, MA 02142, USA), *Olweus' Core Program Against Bullying and Antisocial Behavior: A Teacher Handbook* (Olweus, 2001), the *Revised Olweus Bully/Victim Questionnaire* (Olweus, 1996) with accompanying PC program, and a video cassette on bullying (Olweus & Limber, 1999). More information about the intervention program and ordering of materials can be obtained from Olweus@online.no or nobully@clemsun.edu.

References

- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation*. Chicago: Rand McNally.
- Dishion, T. J., McCord, J., & Poulin, F. (1999). When interventions harm: Peer groups and problem behavior. *American Psychologist*, 54, 755–764.

- Gottfredson, G. D. (1987). Peer group interventions to reduce the risk of delinquent behavior: A selective review and a new evaluation. *Criminology*, 25, 671–714.
- Judd, C. M., & Kenny, D. A. (1981). *Estimating the effects of social interventions*. New York: Cambridge University Press.
- Lipsey, M. W. (1992). Juvenile delinquency treatment: A meta-analytic inquiry into the variability of effects. In T. D. Cook, H. Cooper, D. S. Corday, H. Hartman, L. V. Hedges, R. J. Light, et al. (Eds.), *Meta-analysis for explanation: A casebook* (pp. 83–125). New York: Russell Sage.
- Miller, G. A., & Chapman, J. P. (2001). Misunderstanding analysis of covariance. *Journal of Abnormal Psychology*, 110, 40–48.
- Olweus, D. (1986). *The Olweus Bully/Victim Questionnaire*. Mimeo. Bergen, Norway: Research Centre for Health Promotion, University of Bergen.
- Olweus, D. (1991). Bully/victim problems among schoolchildren: Basic facts and effects of a school based intervention program. In D. Pepler, & K. Rubin (Eds.), *The development and treatment of childhood aggression* (pp. 411–448). Hillsdale, NJ: Erlbaum.
- Olweus, D. (1993). *Bullying at school: What we know and what we can do*. Oxford: Blackwell.
- Olweus, D. (1994a). Annotation: Bullying at school: Basic facts and effects of a school based intervention program. *Journal of Child Psychology and Psychiatry*, 35, 1171–1190.
- Olweus, D. (1994b). Bullying at school: Long-term outcomes for the victims and an effective school based intervention program. In L. R. Huesmann (Ed.), *Aggressive behavior: Current perspectives* (pp. 97–130). New York: Plenum.
- Olweus, D. (1996). *The Revised Olweus Bully/Victim Questionnaire*. Mimeo. Bergen, Norway: Research Centre for Health Promotion, University of Bergen.
- Olweus, D. (1999a). Noen hovedresultater fra Det nye Bergensprosjektet mot mobbing og antisosial atferd. [Some key results from The New Bergen Project against Bullying and Antisocial Behaviour.] Manuskript. HEMIL-senteret, Universitetet i Bergen.
- Olweus, D. (1999b). *Norway*. In P. K. Smith, Y. Morita, J. Junger-Tas, D. Olweus, R. Catalano, & P. Slee (Eds.), *The nature of school bullying: A cross-national perspective* (pp. 28–48). London: Routledge.
- Olweus, D. (2001). *Olweus' core program against bullying and antisocial behavior: A teacher handbook*. Bergen, Norway: Research Centre for Health Promotion (Hemil Center).
- Olweus, D. (2002). *Mobbning i skolan: Grundläggande fakta, ett effektivt åtgärdsprogram och en ny nationell satsning i Norge* [Bullying in school: Basic facts, an effective intervention program, and a new national initiative in Norway]. In I. E. Menckel (Ed.), *Vänskap vinner: Om mobbning och kränkande behandling i skolan* (pp. 27–59). Halmstad, Sverige: Högskolan i Halmstad.
- Olweus, D., & Alsaker, F. D. (1991). Assessing change in a cohort longitudinal study with hierarchical data. In D. Magnusson, L. R. Bergman, G. Rudinger, & B. Törestad (Eds.), *Problems and methods in longitudinal research* (pp. 107–132). New York: Cambridge University Press.
- Olweus, D., & Limber, S. (1999). *Blueprints for violence prevention: Bullying Prevention Program*. Institute of Behavioral Science, University of Colorado, Boulder, USA.
- Porter, A. C., & Raudenbush, S. W. (1987). Analysis of covariance: Its model and use in psychological research. *Journal of Counseling Psychology*, 34, 383–392.
- Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd edn). Newbury Park, CA: Sage.
- Roland, E. (1989). Bullying: The Scandinavian research tradition. In D. P. Tattum, & D. A. Lane (Eds.), *Bullying in schools* (pp. 21–32). Stoke-on-Trent, UK: Trentham Books.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental design for generalized causal inference*. Boston: Houghton-Mifflin.
- Smith, P. K., & Sharp, S. (1994). *School bullying: Insights and perspectives*. London: Routledge.
- Smith, P. K., Madsen, K. C., & Moody, J. C. (1999). What causes the age decline in reports of being bullied at school? Towards a developmental analysis of being bullied. *Educational Research*, 41, 276–285.
- Smith, P. K., Pepler, D., & Rigby, K. (2004). (Eds.) *Bullying in schools: How successful can interventions be?* Cambridge: Cambridge University Press.
- Solberg, M., & Olweus, D. (2003). Prevalence estimation of school bullying with the Olweus Bully/Victim Questionnaire. *Aggressive Behavior*, 29, 239–268.
- Stevens, V., de Bourdeaudhuij, I., & Van Oost, P. (2000). Bullying in Flemish schools: An evaluation of anti-bullying intervention in primary and secondary schools. *British Journal of Educational Psychology*, 70, 195–210.
- Tikkanen, T., & Junge, A. (2004). Realiserings av en visjon om et mobbefritt oppvekstmiljø for barn og unge. Stavanger, Norway: Rogalandsforskning

Copyright of Psychology, Crime & Law is the property of Taylor & Francis Ltd. The copyright in an individual article may be maintained by the author in certain cases. Content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.