Prevalence Estimation of School Bullying With
the Olweus Bully/Victim Questionnaire

Mona E. Solberg* and Dan Olweus*

Research Center for Health Promotion, University of Bergen, Bergen, Norway

The key aim of the present research was to study the “functionality” of two global variables in the Olweus Bully/Victim Questionnaire and to examine the appropriateness of different cutoff points of these variables for prevalence estimation. Several empirical and conceptual analyses strongly attested to the functionality of the two selected variables in terms of construct validity and selected measurement properties. Similarly, a number of analyses indicated that (having been bullied/having bullied other students) “2 or 3 times a month” was a reasonable and useful lower-bound cutoff point. With this cutoff point, “involved” students, victims, and bullies differed very markedly and in clearly different ways from “non-involved” students in conceptually related variables. Prevalence estimates derived in this way can be conveniently obtained, have a reasonably well-defined meaning, can be easily understood by users, and can be reproduced unambiguously by different researchers/administrators and at different times. An important background for the article is the fact that several common methods, including peer nominations, are not well suited for prevalence estimation. Prevalence data for victims, bullies, and bully-victims are also presented. All data were derived from the New Bergen Project Against Bullying, comprising a sample of 5,171 students from 37 schools in the town community of Bergen, Norway. At the time of the data collection, the spring of 1997, the 2,544 girls and 2,627 boys were in grades 5 through 9, with modal ages of 11 through 15 years. Aggr. Behav. 29:239–268, 2003.

Keywords: prevalence; victims; bullies; bully-victims; cutoff point; psychosocial adjustment

INTRODUCTION

In epidemiology, prevalence usually refers to the number of persons with a defined disease or condition existing at a particular point in time (point prevalence) or within a specified time period (period prevalence or cumulative prevalence) relative to the total number of persons in the group or population “exposed to risk” [see, e.g., Hill, 1977; Last, 1983; Olweus, 1989]. If clearly defined and operationalized, a prevalence estimate reflecting the presence/absence of a
disease, characteristic, or condition is a very useful measure that can be employed to make comparisons among groups from different geographical areas or cultures or with different special characteristics of interest. Such an estimate also permits comparisons between different time periods to assess secular changes or the situation before and after an intervention. Although prevalence is a concept used to characterize groups or populations, it is necessary to collect information at the individual level to generate such an aggregate estimate.

Translated into the area of bully/victim problems in school, a period prevalence estimate of victimization, “having been bullied” or “being a victim,” refers to the proportion or percentage of students in a school or other meaningful unit who have been exposed to bullying/victimizing behavior by other students with some defined frequency within a specified time period in the group of interest. A period prevalence estimate of bullying, “having bullied other students” or “being a bully,” can be defined in a parallel way as the proportion or percentage of students who, within the specified time period, have exposed one or more other students to bullying/victimizing behavior with some defined frequency.

This relatively strict definition of prevalence is closely linked to the use of the term in epidemiology. However, many of the “prevalence estimates” or relative frequency estimates reported in the literature on school bullying deviate in one or several respects from such a definition. Poor adherence to the basic meaning of the concept of prevalence in this area is no doubt a major reason for the considerable variability in bully/victim estimates presented in the literature [see e.g., Farrington, 1993; Olweus, 1993; Schuster, 1996; Schwartz et al., 2001; Smith et al., 1999b], over and above the real differences that may exist among different school populations, cultural groups, or countries.

The variability in prevalence rates observed is related to a number of factors, several of which will be briefly discussed. First, studies reporting prevalence rates of bully/victim problems have relied on different data sources, such as peer and teacher nominations or ratings and self-report questionnaires. As will be argued herein, some data sources are less well suited for prevalence estimation than others. Second, some researchers provide their participants with a definition or explanation of what is meant by bullying, and others do not. Without a clear definition, the individual participant who is to respond to a questionnaire or make ratings is given more room for subjective interpretation of what is meant by bullying. This will of course increase variability.

Third, studies may differ with regard to the “reference period” or time frame used in measuring bullying. The time period referred to can, for example, be a whole (school) year, one (school) term, or the past 2 or 3 months. In other studies, prevalence estimates only refer to the current situation, or no information at all is provided about the time period that the participants should relate their assessment to. Fourth, response and rating categories may vary in both number and specificity. Such categories may consist of simple yes-no dichotomies, of various applicability categories such as “does not apply at all” or “applies exactly,” or of relatively vague frequency alternatives varying from “seldom” to “very often” or from more specific temporal categories such as “not at all in the past couple of months” to “several times a week.”

Fifth, some studies base their prevalence estimates on a single item/variable, whereas others use some form of composite score or scale index consisting of, for example, the mean or sum of several variables/items or ratings. As will be shown in more detail herein, there are some problems associated with use of composite or scale scores for prevalence
estimation. Finally, studies use different thresholds or criteria for differentiating victims from non-victims and bullies from non-bullies.

The two latter problems, in particular, will be elaborated by more detailed reference to two of the most commonly used methods for prevalence estimation in the field: peer nominations and self-report questionnaires.

**Peer Nominations/Ratings and Self-Report Questionnaires**

In studies using peer nominations, the students are typically asked to nominate a fixed (often three students of own gender) or optional number of classmates who fit each of several behavioral descriptions (e.g., is often picked on by other kids, bullies other students a lot). The nominations received from peers for a given item/descriptor are then summed and often standardized within classrooms (which eliminates between-classroom variation). If several items are used to form a scale, the relevant standardized nominations are combined into a mean or sum score. To obtain prevalence estimates from these aggregate scores, investigators choose some cutoff point on the distribution of scores, such as 1 standard deviation or .75 standard deviation units above the mean: students at and above this cutoff point are defined as victims (or bullies), whereas those below the chosen cutoff point are classified as non-victims (or non-bullies). If only a single item/descriptor is used to identify victims or bullies, a similar procedure may be used. Another possibility is to decide that at least a certain proportion, or possibly a certain number, of classmates must have nominated a student for a particular descriptor to classify him or her as a victim or a bully.

Peer ratings differ in some ways from peer nominations. A whole class or a smaller group of selected students make assessments of their classmates on certain rating scales with a number of fixed scale categories varying from, for example, “seldom” to “very often” or from “a little” to “a lot.” From a prevalence perspective, however, the problems with this technique are basically the same as with nominations.

An important problem with such approaches relates to the fact that the procedures used to arrive at a cutoff point or differentiating criterion are usually quite complex, difficult to reproduce, and more or less arbitrary. The prevalence estimates arrived at are likely to depend on an array of factors related to the number of students in the classroom, the variations in level of problems among classrooms/schools, if the number of choices are fixed or free, and if, and how, such nominations are standardized, and so on. All of this will make it difficult, if not impossible, for different researchers to “reproduce” the differentiating criterion employed in a particular study and arrive at prevalence estimates that have basically the same meaning. And if this cannot be achieved, meaningful comparisons across groups and time are not possible or are severely limited. In addition, use of composite scores, in particular if they are based on a number of complex operations, as in these cases, typically do not have a clear meaning or conceptual interpretation. How can one specify in a precise, meaningful, and reproducible way the criterion/criteria actually used in these approaches to differentiate victims from non-victims and bullies from non-bullies?

In addition, the decision rules used in choosing a relevant cutoff point often seem to be ad hoc and fairly arbitrary. Why should the researcher use a distance of, for example, 1 standard deviation rather than a .75 or .67 standard deviation above the mean as a cutoff point? Or why should the researcher require that, say, 20% rather than 35% of the classmates have made certain nominations to classify a peer as a victim? There may exist some statistical/psychometric and possibly substantive considerations behind such decisions, but from a
prevalence perspective, the end result is likely to be quite different depending on which choices are made. Often, the rationale for choosing one alternative rather than another is not discussed at all.

In this article, we emphasize that the arguments just presented are derived from a prevalence estimation perspective. It is important to point out that these arguments in no way imply a general devaluation of peer nomination or rating techniques. On the contrary, such methods may be quite useful for a number of purposes where good prevalence estimation is not a key issue. One example is when the aim of the study is rough identification of groups of relatively “extreme” victims and bullies, in addition to non-involved students, whose characteristics are to be examined and compared. Examples of various kinds of peer nomination/peer rating techniques, which have been used in the context of bully/victim problems, are Boulton, 1995; Boulton and Smith, 1994; Boulton and Underwood, 1992; Björkqvist et al. 1982; Crick et al., 2001; Graham and Juvonen, 2001; Juvonen et al., 2001; Lagerspetz et al., 1982; Lowenstein, 1978; Olweus, 1973, 1978; Pellegrini, 2001; Pellegrini and Bartini, 2000; Perry et al., 1988, 1990; Salmivalli et al., 1996a, 1996b, 1998; Schuster, 1999; Schwartz, 2000; and Stephenson and Smith, 1989. Studies such as these have provided useful information on different aspects of bully/victim problems. However, the peer nomination or rating procedures that have been employed are, in our view, not well suited for estimation of prevalence or relative frequency, as outlined previously [also see, e.g., Smith et al., 1999a, p 270, for a similar statement].

Several of the problems just discussed are avoided when a sum/mean composite score derived from a self-report questionnaire is used for prevalence estimation because with such an instrument, component variables or items can be combined into a composite scale or index in a simple and relatively unambiguous way. The result is that a chosen cutoff point can be more easily identified and reproduced at a later time or by other researchers. Accordingly, comparisons of prevalence rates in different groups and across time are likely to be more meaningful. Nonetheless, a general problem with composite scores is that they do not have a quite unambiguous meaning since such a score can be obtained in a number of different ways. This also means that a prevalence estimate based on a composite score (indicating “presence of a certain condition”) is typically somewhat more abstract and general than an estimate derived from a single variable.

In sum, all of the previous considerations seem to suggest that from the perspective of interpretability or meaning, and reproducibility, use of a single variable/item with quite specific response alternatives is the “method of choice” for prevalence estimation. Although very important, interpretability and reproducibility are not the only concerns in prevalence estimation. A chosen variable used for prevalence estimation must also be shown to “function reasonably well” for its intended purposes and to have desirable psychometric properties. What this may mean in the present context and more generally is specified in more detail herein.

To escape as much as possible the problems discussed previously, the second author developed his Bully/Victim Questionnaire [Olweus, 1986, 1996]. This questionnaire, which is filled out anonymously by the students, contains a detailed definition or explanation of bullying (see later herein) read aloud to the students by the person administering the survey. It specifies a clear time frame or “reference period” that is thought to constitute a natural (and not too long) memory unit for the students (“in the past couple of months,” approximately representing the time period from the beginning of school after the summer or Christmas vacation up to the day of the survey). Also, most questions have a clear spatial
reference, asking about events and activities having occurred “at school.” The response alternatives to the key questions are quite specific: “I haven’t been bullied/bullied other students at school in the past couple of months,” “only once or twice,” “2 or 3 times a month,” “about once a week,” and “several times a week,” usually coded on a 5-point scale from 0 to 4 or 1 to 5.

The definition of bullying is followed by the “global” question: “How often have you been bullied at school in the past couple of months?” together with the five response alternatives listed previously. Somewhat later in the Questionnaire, a parallel question is asked about bullying other students: “How often have you taken part in bullying another student(s) at school in the past couple of months?” These two global measures are the key items or variables used to make prevalence estimates in the present article. The two global questions are each followed by seven more specific questions about how often particular (verbal, physical, indirect, racist, etc.) forms of bullying have occurred (in the newest version of the Questionnaire, an eighth question about sexual bullying has been added).

Examples of studies that have used earlier versions of the Olweus Bully/Victim Questionnaire are Nansel et al., 2001; Olweus, 1993; O’Moore et al., 1997; Pellegrini et al., 1999; Smith and Sharp, 1994; Smith et al., 1999b; Whitney and Smith, 1993. Studies that have used other self-report questionnaires include Alasker and Brunner, 1999; Arora, 1994; Arora and Thompson, 1987; Austin and Joseph, 1996; Berthold and Hoover, 2000; Brunner, 1999; Mynard and Joseph, 2000; Rigby, 1996, 1997.

Overview of Article

To examine the “functionality” of the two global variables, we relate them to several variables with which they can be expected to be positively associated, on conceptual grounds and/or on the basis of previous research. Three scales of internalizing problems (perceived social disintegration, depressive tendencies, and global negative self-evaluations/poor self-esteem) in the case of “being bullied”) and two scales of externalizing problems (aggression and antisocial behavior in the case of “bullying other students”) are being used. If the expected pattern of results emerges from these analyses, this will attest to the construct validity of the key variables. For convenience, we sometimes use the general term “psychosocial adjustment” to refer to problems or variables of either an internalizing or an externalizing nature.

A number of studies have documented that victims of bullying experience various kinds of internalizing problems [for a review, see Hawker and Boulton, 2000]. Examples of publications that have reported the expected relationships for internalizing problems, such as social isolation or loneliness, low self-esteem, and depressive tendencies, are Alsaker and Olweus, 1986; Austin and Joseph, 1996; Björkqvist et al., 1982; Boivin and Hymel, 1997; Boivin et al., 1995; Boulton and Smith, 1994; Callaghan and Joseph, 1995; Crick and

1In earlier versions of the Olweus Bully/Victim Questionnaire, the response alternatives “sometimes” and “now and then” have been used. In the revised version, they have been replaced by the more precise term “2 or 3 times a month.”

2In the original version of the Olweus Bully/Victim Questionnaire, “this term” was used as a reference period. To avoid different lengths of the reference period due to varying definitions of “term” in different school systems, “this term” was replaced by the expression “in the past couple of months” in the revised version of the Questionnaire [Olweus, 1996]. In the New Bergen Project Against Bullying reported on in this article, however, the data were collected in May-June (of 1997), and the most natural memory unit for the students was thought to be “this term” representing the 4- to 5-month period from the Christmas vacation up to the time of the survey.

Studies reporting a relative excess of externalizing problems, such as aggression and antisocial behavior among bullies, include the following: Baldry and Farrington, 2000; Berthold and Hoover, 2000; Farrington, 1992, 1993; Haynie et al., 2001; Junger-Tas and Van Kesteren, 1999; Kumpulainen et al., 1998; Lagerspetz, et al., 1982; Nansel, et al., 2001; Olweus, 1973, 1978, 1994; Pellegrini et al., 1999; Rigby and Cox, 1996; and Stephenson and Smith, 1989.

To explore the possibility that obtained relationships are mainly or to a considerable degree a function of “shared method variance” (possibly arising from the fact that both sets of variables are measured with self-reports), we also report the “cross-over” correlations between being bullied and the externalizing scales on one hand and between bullying other students and the three internalizing scales on the other. If the cross-over correlations are considerably lower than the predicted relationships and in particular close to zero, this will attest to another aspect of construct validity, usually referred to as discriminant validity [e.g., Campbell and Fiske, 1959; Nunnally and Bernstein, 1994].

Since, in a scientific context, bullying is generally defined as something that occurs over time and with some repetitiveness [e.g., Olweus, 1978, 1993, 1999a], we also examine the relationship between being bullied and a question about how long the bullying has lasted.

In addition, we report the correlations between prevalence estimates based on global and specific measures of bullying. This is done at the school level because prevalence is basically a characteristic of a larger unit, and it is important to be able to make reliable differentiations among schools. Since both global and specific measures aim to capture roughly the same phenomenon, we expect a fairly high correlation between them, and this correlation can be loosely interpreted as a kind of reliability coefficient. A relatively high correlation would certainly be a desirable psychometric property of the variable(s) concerned.

The analyses just referred to are used in the important context of choosing a suitable cutoff point for classifying a student as a victim/non-victim or a bully/non-bully. Our point of departure is that the response category (having been bullied/having bullied other students) “2 or 3 times a month” (in the past couple of months) is a meaningful and useful lower-bound value or cutoff point for this classification: students who have been bullied/bullied other students “2 or 3 times a month” or more often are classified as victims or bullies, respectively, whereas those who respond that they have not been bullied/not bullied other students or have been bullied/have bullied other students “only once or twice” are considered non-victims or non-bullies.

This view of what constitutes a suitable cutoff point is based on both conceptual arguments—that bullying is defined as something that occurs over time and with a repetitive nature—and tradition—the cutoff point of “2 or 3 times a month” (or the roughly corresponding terms “now and then” or “sometimes” in earlier versions of the Questionnaire) has been used in a number of large-scale studies [see, e.g., Olweus, 1993; O’Moore et al., 1997; Smith and Sharp, 1994; Smith et al., 1999b]. In addition, this cutoff point has been employed in the context of assessments for practical intervention purposes [e.g., Olweus and Limber, 1999]. The traditional use of this cutoff point has also been based, to some extent, on what may be called strategic reasons (discussed later in this article).
In the present context, the appropriateness and meaningfulness of the traditional cutoff point is explored in considerable detail. In addition, we examine the possible usefulness of two alternative cutoff points, that between the categories of “not (been) bullied” and “only once or twice” and that between “2 or 3 times a week” and “about once a week.” It should be generally emphasized that the issue of choosing a suitable cutoff point or threshold value has not received much attention in the literature on bully/victim problems.

Finally, this article gives a brief presentation of prevalence data for victims and bullies (including bully-victims, victims only, and bullies only) derived from approximately 5,000 students in the New Bergen Project Against Bullying from 1997 [Olweus, 1999b; Olweus and Limber, 1999]. However, these data are given relatively little detailed attention in this article because the main focus here is on the more general and basic issues involved in prevalence estimation rather than on the particular empirical results obtained.

METHODS

The data that form the basis of this study were collected in the town community of Bergen, Norway, in the spring of 1997 (May-June) as part of a large-scale cohort-longitudinal intervention project under the direction of the second author [see Olweus, 1999b; Olweus and Limber, 1999].

Participants

The target population of this study was restricted to students in grades 5 through 9, with modal ages of 11 through 15 years, in Bergen in the school year of 1996/97. Because it is very uncommon to be held back in school in Norway, there were only small variations in age among students who belonged to the same class or grade level. Grades 5 through 7 are the three highest grades in elementary/primary school, and grades 8 and 9 are the two first grades of lower secondary school.

For reasons of design, only schools with at least two parallel classes at each relevant grade level could be included in the study. A total of 54 schools were invited to participate in the investigation. The final sample consisted of 37 schools (68.5%): 26 elementary schools, 9 lower secondary schools, and 2 “combined” schools (with both a primary and a secondary level), with a total of 259 classes participating in the study (177 at the elementary level and 82 at the lower secondary level).

The total number of potentially available students in the 259 classes was 5,825. We were finally able to get valid data from 88.8% of the total population, or 5,171 students (2,544 girls and 2,627 boys). From all these students, we had obtained written positive consent from both the individual student and his or her parent/caregiver.

Although the target population of this study only comprises schools in Bergen, there are grounds for assuming, on the basis of earlier large-scale studies [e.g., Olweus, 1991, 1993], that the findings on bully/victim problems are at least roughly representative of other town communities in Norway.

Procedure

The data were collected at school in a 2-hour session with a break in the middle. In the break, the students stayed in their classroom and orange juice was served. Two specially trained research assistants administered the large questionnaire, which was composed of
several sub-questionnaires, to the students in their ordinary classrooms. Teachers were absent when the students filled out their questionnaires. Detailed instructions on how to respond and an explanation or “definition” of what is meant by bullying was included in the questionnaire. In addition, the instructions were read aloud by the research assistants. The students were strongly encouraged to give sincere answers. They were told that nobody at school or at home would be informed about how they responded. An abbreviated version of the questionnaire was used with students in grades 5 and 6.

Measures

The first version of the Olweus Bully/Victim Questionnaire was developed by the second author in connection with a nationwide campaign against bullying in Norway in 1983 [Olweus, 1986, 1993]. A revised version of this questionnaire was used in the present study [Olweus, 1996]. The Revised Olweus Bully/Victim Questionnaire contains 36 main questions on various aspects of bully/victim problems and a couple of questions covering related areas. For a few of the main questions, there are several sub-questions.

The explanation of bullying, which was included in the questionnaire, was intended to capture all three main elements of the definition of bullying: the intention to harm the victim, the repetitive nature of bullying, and the imbalance in power between the victim and the perpetrator(s). In the revised version of the Questionnaire, the explanation of bullying was slightly expanded to include several more specific forms of bullying. In addition, it was made clear when teasing should and should not be considered bullying. The “definition” presented to the students reads as follows [Olweus, 1996]:

“We say a student is being bullied when another student or several other students

- say mean and hurtful things or make fun of him or her or call him or her mean and hurtful names
- completely ignore or exclude him or her from their group of friends or leave him or her out of things on purpose
- hit, kick, push, shove around, or threaten him or her
- tell lies or spread false rumors about him or her or send mean notes and try to make other students dislike him or her
- and do other hurtful things like that.

These things may take place frequently, and it is difficult for the student being bullied to defend himself or herself. It is also bullying when a student is teased repeatedly in a mean and hurtful way.

But we don’t call it bullying when the teasing is done in a friendly and playful way. Also, it is not bullying when two students of about the same strength or power argue or fight.”

The global measures of being bullied and bullying other students. The two global measures were briefly described in the first part of this article (see also Footnote 1). Two versions of the global measures were used in the present study: (1) a dichotomized version in which students who had been bullied/bullied others “2 or 3 times a month” or more often (this term) were classified as victims/bullies, and students who had not been bullied/bullied others (this term) or replied “only once or twice” were categorized as non-victims/non-bullies and (2) a version with four graded categories in which the two highest of the five response categories (”about once a week” and ”several times a week”) were collapsed.
A specific measure of being bullied/bullying other students. A dichotomous index was constructed based on the seven specific questions about the various forms of victimization. To be denoted a victim on the specific measure, a student had to have been bullied “2 or 3 times a month” or more often in at least one of the seven different ways specified in the Questionnaire. The various ways of victimization concerned, in abbreviated form, being bullied verbally, being excluded from a group or completely ignored, being bullied physically, having false rumors spread, having money and other things taken away or damaged, being threatened or forced to do things, and being bullied about one’s race or color. The five response categories to each of these specific questions were the same as for the global questions previously. A corresponding specific measure of bullying other students (categorizing a student as a bully or a non-bully on the specific measure) was constructed on the basis of the seven different ways of bullying other students.

A combined measure of being bullied/bullying other students. Students who reported that they both had been bullied and had bullied other students “2 or 3 times a month” or more often were identified by combining the two global measures. Combining these two measures, we got four groups of students: (1) those who were both victims and bullies (bully-victims), (2) those who were victims only (pure victims), (3) those who were bullies only (pure bullies), and (4) those who were neither victims nor bullies (not involved). To be classified as a bully-victim or placed in any of the other categories, valid (non-missing) responses on both global measures were necessary.

Duration of being bullied. The duration of victimization was assessed by the question, “How long have you been bullied?” The six response categories to this question were as follows: I haven’t been bullied at school this term (coded 0), it lasted less than a week (coded 1), it lasted about a month (coded 2), it has lasted all term (coded 3), it has lasted about a year (coded 4), and it has gone on for several years (coded 5).

Social disintegration in class/peer group. To assess the students’ degree of perceived social disintegration, we used the mean score of four items. One item concerned the extent to which the students felt accepted among classmates (“I feel less well liked than other students in my class”), and three other items reflected the degree to which they experienced feelings of not belonging when being with classmates or peers in general (e.g., “When I am together with other children/young people, I think I don’t quite belong there”). Items expressing feelings of not belonging have been denoted social alienation by some researchers [e.g., Alsaker, 2000; Alsaker and Flammer, 1996]. There were six response alternatives to these statements: seldom or never (coded 1), once or twice (coded 2), sometimes (coded 3), fairly often (coded 4), often (coded 5), and very often (coded 6). A high score thus indicated a high level of self-perceived social disintegration. Cronbach’s alpha for this variable was .76.

Global negative self-evaluations. To assess the degree to which students had generally negative views of themselves, we used the mean score of six items, largely based on a scale described by Alsaker and Olweus [1986]. The items were formulated as statements (e.g., “I have often wanted to be someone else” and “I feel quite often that I am a failure”), and the students had to report the degree to which these statements applied to them: doesn’t apply at all (coded 1), doesn’t apply well (coded 2), applies somewhat (coded 3), applies fairly well (coded 4), applies well (coded 5), and applies exactly (coded 6). A high score thus indicated a high level of global negative self-evaluations. The alpha of this scale was .84.

Depressive tendencies. To assess the degree of depressive tendencies in students, we used the mean scores on a seven-item scale (alpha = .78) developed by Olweus and Alsaker [see Alsaker et al., 1991]. The wording of the different items (e.g., “I am often sad without
seeing any reason for it” and “I think my life is mostly miserable”) is also reported in Holsen et al. [2000]. Owing to the young age of students in the grade 5 and 6 cohorts, the item “Sometimes I think my life is not worth living” was replaced by the item “I don’t think I have anything to look forward to.” The response alternatives were the same as for global negative self-evaluations (see previously). A high score indicated a high level of self-reported depressive tendencies.

**General aggression.** The general aggression scale was developed by the second author to assess the degree to which students used mainly physically and verbally aggressive behavior in their relations with peers and teachers at school (e.g., “If I disagree with a boy or a girl my age, I easily get angry and yell at him or her,” “I fight fairly often with other students at school,” and “It is quite all right to make trouble for a teacher whom you don’t like”). The response alternatives were the same as for global negative self-evaluations (see previously). The scale was based on nine items, and Cronbach’s alpha was .84. A high score indicated a high level of general aggression. Only a selection of these items were included in the abbreviated version of the Questionnaire for grades 5 and 6, and, accordingly, only data from the three oldest cohorts are used in the present article.

**Antisocial behavior.** To assess the degree to which students were involved in antisocial activities, we used the 17-item Bergen Questionnaire on Antisocial Behavior developed by Olweus and described in detail in Bendixen and Olweus [1999]. Ten of these items concern relatively non-serious, high-prevalence activities (e.g., avoid paying for things, scribble on the school building, and skip school a whole day), whereas seven items cover more serious, low-prevalence activities (e.g., sign someone else’s name to get money or other things; purposely break chairs, tables, etc., in school; and break into a shop, etc.). The reliability of the total scale was .84.

**Missing data.** To understand the pattern of missing data that emerged on closer analysis, it is important to know that the questionnaire was quite comprehensive. And, as mentioned, the students in the two lowest grades had a shorter form of the instrument (but the same general sequence of questions as for the higher grades). Furthermore, the questionnaire was divided into two parts, and students who had not completed questions toward the end of the first part before the short break were told to skip these questions and to start with the questions at the beginning of the second part after the break. The questions about bullying were placed at the beginning of the first part of the questionnaire, and the questions about antisocial behavior at the beginning of the second part.

For these two blocks of questions, there were generally few missing data, some 2% to 3% for the global questions on bullying and 1.5% for antisocial behavior. In contrast, for the questions on social disintegration, global negative self-evaluations, depressive tendencies, and general aggression that were located toward the end of their respective parts, the percentage of missing data was higher, varying between 7.8% and 13.2%.

In addition, there was a clear trend for students in lower grades, both overall and within each grade group (5-6 and 7-9), to have more missing data than older students. Finally, there was a similar trend for students with (self-reported) reading difficulties compared with students without such problems. In sum, students with less developed reading skills had relatively more missing data and, quite naturally, this was particularly marked with regard to questions that were located toward the end of their respective parts.

We conducted several analyses to check for possible biasing effects of this pattern of missing data, e.g., by restricting the statistical analyses to students in the four highest grades. Analyses were also carried out separately for grade 9 students who had the lowest levels of
missing data. The pattern of findings was generally very much the same as for the unrestricted analyses reported in the article, but in some cases the correlations and effect sizes were somewhat larger (as would be expected). Accordingly, if there was an effect of the missing data, it was likely to be one of slight underestimation.

**Statistical Analyses**

A major strategy of the present article was to relate different frequencies of bullying, as indicated by the various response alternatives to the two global questions, to conceptually related variables. To somewhat simplify the analyses and results, we collapsed the two highest response alternatives into one category, as mentioned earlier. Accordingly, for these analyses, we had four groups of subjects representing graded levels of “being bullied” and “bullying other students.” These frequency groups or graded categories were named Group 0 (“not been bullied” and “not bullying other students”), Group 1 (“only once or twice”), Group 2 (“2 or 3 times a month”), and Group 3 (“about once a week” or more often).

For the major analyses, the two higher-level and the two lower-level groups were combined into Group 2&3 and Group 0&1, respectively. These groups represent what has traditionally been designated as victims/bullies and non-victims/non-bullies. To facilitate reading of this article and in agreement with our point of departure stated in the “Overview of Article” section, we will use these terms for these groups in several analyses. They will, however, be put within quotation marks (“victims” and “bullies”) to signify that we have not yet reached a firm conclusion on what may be the most useful cutoff point. It should be noted that construction of the groups was made quite independently for the two global questions about being bullied and bullying other students.

The comparisons of Group 2&3 vs. Group 0&1 were followed up by several other analyses. We were particularly interested in comparing Group 1 (“only once or twice”) and Group 2 (“2 or 3 times a month”). In these analyses, we were able to examine the extent to which the traditional cutoff point was able to differentiate between the two groups on either side of the dividing line. These comparisons represent a stronger test of the traditional choice of cutoff point than the analyses consisting of Group 2&3 vs. Group 0&1.

Also, comparisons between Group 1 and Group 0 and between Group 3 and Group 2 were of considerable interest. These comparisons allowed us to explore possible differences between adjacent frequency categories, and results were of importance in evaluating possible alternative cutoff points.

Finally, we tested the linearity of the relationship between the four levels of victimization/bullying and the conceptually related variables. For that purpose, we conducted trend analyses with polynomial coefficients in one-way ANOVAs. In that context, we also reported (point-biserial) correlations between the two global victimization/bullying variables and the conceptually related variables as well as the previously mentioned “cross-over” correlations to examine the discriminant validity of the global variables.

With 3 degrees of freedom, only three orthogonal contrasts could be constructed [e.g., Kirk, 1982]. For the orthogonal contrasts (Group 2&3 vs. Group 0&1, Group 1 vs. Group 0, and Group 3 vs. Group 2), we set alpha at .05 for each comparison and used one-tailed t-tests, because the direction of the difference was predicted in all cases: higher values on the “dependent variable” for higher-level groups. These predictions were based on both conceptual grounds and on previous empirical research.

The other comparisons (Group 2 vs. Group 1) and the linearity tests were not orthogonal to the ones already specified. We therefore used ordinary Bonferroni adjustment of the alpha
level for these analyses (alpha divided by the number of comparisons). With a one-tailed test and an overall ("familywise") alpha level of .05, alpha was set at .025 for each of these tests.

Generally, when the assumption about equal variances in the two groups compared was not met (at an alpha level of .01), we used the "unequal variance" procedure in SPSS with corrected degrees of freedom.

To give a more concrete impression of the size of the differences, we also report some effect sizes in the form of Cohen’s $d$, which is the difference between the group means divided by the pooled standard deviation [e.g., Cohen, 1988; Rosenthal, 1991].

**RESULTS**

**Being Bullied**

Comparing Group 2&3 and Group 0&1, "victims" and "non-victims" (according to traditional usage). There were large and highly significant differences between the two groups on all three dependent variables examined (Table I). Very similar results were found when data were analyzed for girls and boys separately. "Victims" reported clearly higher levels of social disintegration, global negative self-evaluations, and depressive tendencies than "non-victims". Effect sizes were quite substantial, with $d$-values varying between 1.05 for social disintegration and 0.62 for global negative self-evaluations, for girls and boys combined (effect sizes not given in the table).

**TABLE I. Differences on Scales of Internalizing Problems Between Group 2&3 ("Victims") and Group 0&1 ("Non-victims"), Girls and Boys Combined and Separately**

<table>
<thead>
<tr>
<th>Scale</th>
<th>Girls and boys</th>
<th>Girls</th>
<th>Boys</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Group 0&amp;1 vs. Group 2&amp;3</td>
<td>Group 0&amp;1 vs. Group 2&amp;3</td>
<td>Group 0&amp;1 vs. Group 2&amp;3</td>
</tr>
<tr>
<td>Social disintegration</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>1.95</td>
<td>2.94</td>
<td>1.97</td>
</tr>
<tr>
<td>SD</td>
<td>0.90</td>
<td>1.25</td>
<td>0.92</td>
</tr>
<tr>
<td>n</td>
<td>4174</td>
<td>452</td>
<td>2114</td>
</tr>
<tr>
<td>$t$</td>
<td>16.35***</td>
<td>12.28***</td>
<td>11.09***</td>
</tr>
<tr>
<td>df$^b$</td>
<td>503</td>
<td>221</td>
<td>249</td>
</tr>
<tr>
<td>Global negative self-evaluations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>2.38</td>
<td>3.05</td>
<td>2.64</td>
</tr>
<tr>
<td>SD</td>
<td>1.04</td>
<td>1.34</td>
<td>1.11</td>
</tr>
<tr>
<td>n</td>
<td>4159</td>
<td>446</td>
<td>2104</td>
</tr>
<tr>
<td>$t$</td>
<td>10.19***</td>
<td>7.31***</td>
<td>8.09***</td>
</tr>
<tr>
<td>df$^b$</td>
<td>504</td>
<td>226</td>
<td>274</td>
</tr>
<tr>
<td>Depressive tendencies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>2.15</td>
<td>2.84</td>
<td>2.31</td>
</tr>
<tr>
<td>SD</td>
<td>0.90</td>
<td>1.15</td>
<td>0.94</td>
</tr>
<tr>
<td>n</td>
<td>3955</td>
<td>403</td>
<td>2000</td>
</tr>
<tr>
<td>$t$</td>
<td>11.59***</td>
<td>9.48***</td>
<td>7.96***</td>
</tr>
<tr>
<td>df$^b$</td>
<td>453</td>
<td>199</td>
<td>257</td>
</tr>
</tbody>
</table>

---

$^1$Group 0&1: “not been bullied” and “only once or twice”; Group 2&3: been bullied “2 or 3 times a month” or more often.

$^a$Alpha level was set at .05 for all comparisons.

$^b$The df’s listed do not equal $n_1+n_2-2$ because of correction due to unequal variances.

$^{***p < .001.}$
Comparing subgroups. Table II reports the results from the three sets of comparisons among students in adjacent categories on the global victimization (being bullied) variable. Again, all differences on all three dependent variables were significant, although the $t$-tests naturally were smaller than in Table I (for girls and boys combined). This was due to the reduced number of subjects in the various Table II comparisons and the fact that only directly adjacent response categories on the “independent” victimization variable were implicated. Highly similar results were obtained when the data were analyzed separately for girls and boys (not given in table).

Summarizing the results for all three psychosocial adjustment variables, students who had been bullied “2 or 3 times a month” had higher scores in the negative direction than students who had been victimized “only once or twice” (Group 2 vs. Group 1). The same was true of students in the “only once or twice” group vs. the “not bullied” group (Group 1 vs. Group 0) and the “about once a week or more” group vs. the “2 or 3 times a month” group (Group 3 vs. Group 2).

Linearity of relationships and discriminant validity. When the four graded response categories for the global victimization variable were related to the mean values on the social disintegration scale, there was a clear linear trend for both girls and boys, $F(1, 2313) = 399.77, P < .001$ for girls and $F(1, 2305) = 298.69, P < .001$ for boys (see Fig. 1).

Linear trends were also obvious for the other two dependent variables. For the global negative self-evaluations scale, $F(1, 2301) = 110.93, P < .001$ for girls and $F(1, 2296) = 166.48$,
P < .001 for boys. For the depressive tendencies scale, corresponding values were $F(1, 2177) = 200.91, P < .001$ for girls and $F(1, 2173) = 126.60, P < .001$ for boys. Highly significant linear trends for all three scales were also found when the data were analyzed for lower (5–7) and higher (8–9) grades separately (not reported).

The dichotomized global variable of being bullied ("victims" and "non-victims") correlated .30, .18, and .20 (point-biserial, $r$), all highly significant at $P < .001$, with the internalizing scales of social disintegration, global negative self-evaluations, and depressive tendencies, respectively. (With the 4-point scale of being bullied, these correlations were .36, .22, and .25.) In comparison, the cross-over correlations between the dichotomized global variable of being bullied and the aggression and antisocial scales were close to zero, .02, and .03.

**Frequency and duration of being bullied.** By cross-tabulating the frequency of being bullied and the duration of the bullying, we got information on how long students in the various frequency groups had been bullied. There was a strong positive relationship between these two variables (Table III). About 82% (14.2% + 16.4% + 51.5%) of the students who had been bullied "about once a week" or more often (Group 3) reported that they had been...
bullied 1 month or longer. Also among the students who responded “2 or 3 times a month” on the global measure (Group 2), a clear majority, 64% (29.9% + 7.1% + 27.0%) had been bullied at least 1 month. As for students who responded “only once or twice” (Group 1), the corresponding percentage was considerably lower, about 25% (10.7% + 4.5% + 9.7%). The correlation ($r$) between frequency and duration of being bullied was .73 for all students. When calculated only for students who had been bullied “2 or 3 times a month” or more often, $r$ was .46 (correlations not given in table).

As much as 43.8% of the students in Group 1 reported that the bullying had lasted less than a week. For the students who reported having been bullied “2 or 3 times a month” (Group 2), the corresponding figure was 27.0%.

On the whole, the responses made by the students were rather consistent. About 97% of the students who responded that they had not been bullied on the global measure (Group 0) also reported that they had not been bullied when asked about the duration of the bullying. However, among students who replied “only once or twice” on the global measure (Group 1), about 31% said that they had not been bullied. Also for Groups 2 and 3, there was some such apparent lack of consistency in responding for 9% of the students in Group 2 and 3.6% of the students in Group 3. Some different interpretations of these findings are included in the “Discussion” section.

### Bullying Other Students

**Comparing Group 2&3 and Group 0&1, “bullies” and “non-bullies” (according to traditional usage).** We found highly significant differences between “bullies” and “non-bullies” (Table IV). The “bullies” reported being much more aggressive as well as clearly more involved in antisocial behavior. For both variables, effect sizes were quite large: 1.12 for general aggression and 1.02 for antisocial behavior for girls and boys combined (effect sizes not given in table).

**Comparing subgroups.** Table V reports the results from the three sets of comparisons among students in adjacent response categories on the global bullying variable. All differences between adjacent categories were significant. The mean differences between the two highest response categories (Group 3 vs. Group 2) were somewhat smaller than between the other groups (Group 1 vs. Group 0 and Group 2 vs. Group 1). Very similar results were obtained when the analyses were conducted separately for girls and boys (not given in table).
TABLE IV. Differences on Scales of Externalizing Problems Between Group 2&3 (“Bullies”) and Group 0&1 (“Non-Bullies”), Girls and Boys Combined and Separately

<table>
<thead>
<tr>
<th>Scale</th>
<th>Girls and boys</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Group 0&amp;1 vs.</td>
<td>Group 0&amp;1 vs.</td>
<td>Group 0&amp;1 vs.</td>
<td>Group 0&amp;1 vs.</td>
<td>Group 0&amp;1 vs.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Group 2&amp;3</td>
<td>Group 2&amp;3</td>
<td>Group 2&amp;3</td>
<td>Group 2&amp;3</td>
<td>Group 2&amp;3</td>
<td></td>
</tr>
<tr>
<td>General aggression b</td>
<td>2.35 3.26 2.24 3.27</td>
<td>2.47 3.26 2.47 3.26</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>0.78 1.05 0.71 1.07</td>
<td>0.84 1.05 0.84 1.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>2252 203 1155 48</td>
<td>1097 155 1097 155</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t</td>
<td>12.02***</td>
<td>6.61***</td>
<td>8.97***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>df</td>
<td>223</td>
<td>49</td>
<td>183</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antisocial behavior d</td>
<td>9.58 27.33 8.70 25.67</td>
<td>10.52 27.88 10.52 27.88</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>14.43 23.81 12.85 24.00</td>
<td>15.88 23.77 15.88 23.77</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>4647 320 2387 79</td>
<td>2260 241 2260 241</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t</td>
<td>13.17***</td>
<td>6.25***</td>
<td>11.08***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>df</td>
<td>335</td>
<td>79</td>
<td>263</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a Group 0&1: “not bullied other students” and “only once or twice”; Group 2&3: bullied other students “2 or 3 times a month” or more often.
bAlpha level was set at .05 for all comparisons.
cData for grades 7-9.
dThe df’s listed do not equal n1 + n2 because of correction due to unequal variances.
eData for grades 5-9.

TABLE V. Differences on Scales of Externalizing Problems Between Various Response Categories (Groups) of Global Measure of Bullying Other Students, Girls and Boys Combined

<table>
<thead>
<tr>
<th>Scale</th>
<th>Group 0 vs.</th>
<th>Group 1 vs.</th>
<th>Group 2 vs.</th>
<th>Group 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Group 0</td>
<td>Group 1</td>
<td>Group 2</td>
<td>Group 3</td>
</tr>
<tr>
<td>General aggression c</td>
<td>2.17 2.70</td>
<td>3.16 3.42</td>
<td>1.72 3.42</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>0.69 0.83</td>
<td>1.02 1.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>1483 769</td>
<td>123 80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t</td>
<td>15.03***</td>
<td>4.77***</td>
<td>1.72*</td>
<td></td>
</tr>
<tr>
<td>df</td>
<td>1323 149</td>
<td>149 201</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antisocial behavior e</td>
<td>6.71 16.53</td>
<td>25.18 30.87</td>
<td>199 201</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>11.21 18.44</td>
<td>22.33 25.77</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>3287 1360</td>
<td>199 201</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t</td>
<td>5.21***</td>
<td>18.30***</td>
<td>2.09*</td>
<td></td>
</tr>
<tr>
<td>df</td>
<td>239 1789</td>
<td>318</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a Group 0: “not bullied other students”; Group 1: bullied other students “only once or twice”; Group 2: “2 or 3 times a month”; Group 3: “about once a week” or more often.
bAlpha level set at .05.
cAlpha level set at .025.
dData for grades 7-9.
eData for grades 5-9.

**p < .01.
***p < .001.
**Linearity of relationships and discriminant validity.** When related to the four graded response categories for the global bullying variable, the mean values of the general aggression scale formed a straight line for girls, $F(1, 1199) = 253.997, P < .001$. Also, for boys, the linear trend was highly significant, $F(1, 1248) = 196.724, P < .001$, but there was a slight (nonsignificant) downward deviation between “2 or 3 times a month” (Group 2) and “about once a week” or more often (Group 3) (see Fig. 2).

Corresponding analyses of the antisocial scale also showed clear linear trends for both girls, $F(1, 2462) = 390.487, P < .001$, and boys, $F(1, 2497) = 436.46, P < .001$. For girls, also the quadratic term, $F(1, 2462) = 15.862, P < .001$, was significant. When the data were analyzed separately for lower and higher grades, we also found highly significant linear trends for both scales (antisocial scale: grades 5-7 vs. grades 8-9; general aggression scale: grade 7 vs. grades 8-9).

The correlation was .28 (point-biserial $r$), highly significant at $P < .001$, between the dichotomized global variable of bullying other students (“bullies” and “non-bullies”) and both scales of externalizing problems (the general aggression and the antisocial behavior scales). (For the 4-point variable of bullying other students, the correlations were .39 and .38). In comparison, the cross-over correlations with the three scales of internalizing problems were only .06, .03, and .06.

![Fig. 2. Mean values of general aggression vs. frequency of bullying other students: girls and boys in grades 7 through 9 (N = 2455).](image-url)
Correlations Between Global and Specific Prevalence Estimates

As mentioned in the “Introduction section” (“Overview of Article”) section, we expected a fairly high school-level correlation between prevalence estimates derived from the dichotomized global variables of victimization (“victims”/“non-victims”) and bullying other students (“bullies”/“non-bullies”) and corresponding estimates based on the specific measures of victimization/bullying. This is because both sets of estimates aim to capture roughly the same phenomena, although in somewhat different ways. Since such prevalence estimates are typically used for somewhat larger units such as a school or a community, we report correlations between the two sets of estimates for all 37 schools, with data for boys and girls combined.

The correlation between the dichotomized global and the specific prevalence estimates for being bullied was .79, and for bullying other students, was .77. Due to the somewhat different nature of the two variables, it may not be natural to interpret this correlation as a conventional internal consistency (alpha) coefficient. Nonetheless, one can apply the Spearman-Brown formula [e.g., Nunnally and Bernstein, 1994] to get a rough (lower-bound) estimate of the extent to which schools can be differentiated from one another on the basis of the global and specific measures of victimization/bullying. In the present study, these “reliability” estimates approach .90, being .88 for being bullied and .87 for bullying other students.

Prevalence estimates of being bullied varied considerably among schools, between 4% and 20% at the primary level (grades 5-7) and between 4% and 17% at the lower secondary level (grades 8 and 9). Regarding bullying other students, corresponding school values were 1% and 12% at the primary level and 6% and 16% at the lower secondary level (estimates not given in table).

Prevalence of Being Bullied/Bullying Other Students

The global measure of being bullied. Table VI presents the detailed distribution of the responses on the various response categories. Table VII shows prevalence estimates using the traditional cutoff point between “only once and twice” and “2 to 3 times a month.” The total number of victims was 506, or 10.1% of all students (Table VII). There were significantly more boys than girls who reported being bullied by other students: 11.1% vs. 9.1% (χ² = 5.69, P < .05).

As has been documented in previous research [e.g., Olweus, 1993; O’Moore et al., 1997; Nansel et al., 2001; Rigby, 1996, 1997; Roland, 1989; Smith et al., 1999a; Whitney and Smith,

| TABLE VI. Frequencies of All Response Categories on Global Measures of Being Bullied and Bullying Other Students, Percentages of All Students, Girls, and Boys |
|----------------------------------|------------------|------------------|------------------|------------------|------------------|
|                                  | n                | Not been bullied/not bullied others | Only once or twice | 2 or 3 times a month | About once a week | Several times a week |
| Being bullied                    |                  |                                |                  |                  |                  |                      |
| All students                     | 5015             | 68.2                           | 21.7             | 4.3              | 3.0              | 2.8                  |
| Girls                            | 2482             | 70.5                           | 20.5             | 3.7              | 2.9              | 2.5                  |
| Boys                             | 2533             | 66.0                           | 22.9             | 4.9              | 3.1              | 3.1                  |
| Bullying other students          |                  |                                |                  |                  |                  |                      |
| All students                     | 5035             | 66.1                           | 27.4             | 4.0              | 1.5              | 1.0                  |
| Girls                            | 2497             | 75.2                           | 21.5             | 2.0              | 0.8              | 0.5                  |
| Boys                             | 2538             | 57.2                           | 33.2             | 6.0              | 2.2              | 1.4                  |
we found a general trend for younger students to report being bullied more frequently than older ones (Table VII). The downward trend was not quite uniform, however, in that the age 14 group showed a slight (possibly random) deviation from the general pattern.

The global measure of bullying other students. Table VI also shows the detailed distribution of the responses on the global variable of bullying other students. With the traditional cutoff point, the total number of bullies was 326, or 6.5% of all students (Table VII). A sex difference was evident across all age groups, with the prevalence of bullying other students being two to three times higher for boys than for girls (9.7% vs. 3.2% for all age groups combined; $\chi^2 = 85.39, P < .001$). In particular for boys, the tendency to bully others increased with age.

The combined measure of victimization and/or bullying. As previously explained, bully-victims were identified by combining the global measures of being bullied and bullying others. When the traditional “2 or 3 times a month” was used as the lower-bound threshold value on both global measures, there were 78 bully-victims. This amounted to 1.6% of all students (Table VII). Overall, there were clearly more male (2.3%) than female (0.9%) bully-victims, and this sex difference was evident at each age level. Almost 75% (57/78) of the bully-victims were actually boys. If expressed as percentages of the total number of victims or bullies, the bully-victims constituted some 16% of the victims [78/(78 + 408)] and some 25% of the bullies [78/(78 + 235)] (Table VII). For boys, the prevalence of bully-victims tended to

### TABLE VII. Prevalence of Victims, Bullies, and Bully-Victims by Age and Sex, Cutoff Point Between “Only Once or Twice” and “2 or 3 Times a Month,” Percentages of Girls and Boys Combined, Girls and Boys Separately, and the Specified Categories

<table>
<thead>
<tr>
<th>Age (grade)</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>11-15</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>n</strong></td>
<td>(5)</td>
<td>(6)</td>
<td>(7)</td>
<td>(8)</td>
<td>(9)</td>
<td>(5-9)</td>
</tr>
<tr>
<td>Victims total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Girls and boys</td>
<td>506</td>
<td>12.3</td>
<td>11.0</td>
<td>9.5</td>
<td>10.0</td>
<td>7.1</td>
</tr>
<tr>
<td>Girls</td>
<td>225</td>
<td>11.5</td>
<td>9.5</td>
<td>8.1</td>
<td>8.4</td>
<td>7.3</td>
</tr>
<tr>
<td>Boys</td>
<td>281</td>
<td>13.2</td>
<td>12.5</td>
<td>10.9</td>
<td>11.7</td>
<td>6.9</td>
</tr>
<tr>
<td>Bullies total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Girls</td>
<td>326</td>
<td>4.6</td>
<td>3.7</td>
<td>6.0</td>
<td>8.8</td>
<td>9.6</td>
</tr>
<tr>
<td>Girls</td>
<td>81</td>
<td>2.2</td>
<td>2.1</td>
<td>3.1</td>
<td>5.2</td>
<td>3.9</td>
</tr>
<tr>
<td>Boys</td>
<td>245</td>
<td>7.0</td>
<td>5.3</td>
<td>9.0</td>
<td>12.3</td>
<td>14.8</td>
</tr>
<tr>
<td>Victims only, bullies only, and bully-victims: girls and boys</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Victims only</td>
<td>408</td>
<td>9.1</td>
<td>9.4</td>
<td>7.9</td>
<td>8.8</td>
<td>6.2</td>
</tr>
<tr>
<td>Bullies only</td>
<td>235</td>
<td>2.1</td>
<td>2.3</td>
<td>4.5</td>
<td>7.1</td>
<td>8.5</td>
</tr>
<tr>
<td>Bully-victims</td>
<td>78</td>
<td>2.5</td>
<td>1.5</td>
<td>1.5</td>
<td>1.4</td>
<td>0.9</td>
</tr>
<tr>
<td>Sub total</td>
<td>721</td>
<td>13.7</td>
<td>13.2</td>
<td>13.9</td>
<td>17.3</td>
<td>15.7</td>
</tr>
<tr>
<td>Victims only, bullies only, and bully-victims: girls</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Victims only</td>
<td>197</td>
<td>9.9</td>
<td>8.5</td>
<td>7.3</td>
<td>7.4</td>
<td>6.9</td>
</tr>
<tr>
<td>Bullies only</td>
<td>58</td>
<td>1.1</td>
<td>1.4</td>
<td>2.4</td>
<td>3.8</td>
<td>3.6</td>
</tr>
<tr>
<td>Bully-victims</td>
<td>21</td>
<td>1.1</td>
<td>0.8</td>
<td>0.9</td>
<td>1.1</td>
<td>0.4</td>
</tr>
<tr>
<td>Sub total</td>
<td>276</td>
<td>12.1</td>
<td>10.7</td>
<td>10.6</td>
<td>12.3</td>
<td>10.9</td>
</tr>
<tr>
<td>Victims only, bullies only, and bully-victims: boys</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Victims only</td>
<td>211</td>
<td>8.3</td>
<td>10.4</td>
<td>8.5</td>
<td>10.2</td>
<td>5.6</td>
</tr>
<tr>
<td>Bullies only</td>
<td>177</td>
<td>3.0</td>
<td>3.2</td>
<td>6.7</td>
<td>10.4</td>
<td>13.0</td>
</tr>
<tr>
<td>Bully-victims</td>
<td>57</td>
<td>4.0</td>
<td>2.2</td>
<td>2.2</td>
<td>1.7</td>
<td>1.4</td>
</tr>
<tr>
<td>Sub total</td>
<td>445</td>
<td>15.3</td>
<td>15.8</td>
<td>17.4</td>
<td>22.3</td>
<td>20.0</td>
</tr>
</tbody>
</table>
decrease with age, while for girls the trend was mainly stable. As evident from these figures, the number/percentage of bully-victims was relatively small, and the correlation between the dichotomized victimization and bullying variables was .13 for girls and boys combined: .21 in grades 5–7 and .05 in grades 8–9.

Combining the two global measures and using the same cutoff point, we could also calculate the prevalence of victims only or pure victims (being bullied “2 or 3 times a month” or more often and bullying other students less than “2 or 3 times a month”) and bullies only or pure bullies (correspondingly defined). The percentage of pure victims in the total group was 8.3 and the percentage of pure bullies was 4.8 (girls and boys combined) (Table VII).

Adding together victims only, bully-victims, and bullies only provides information on the total percentages of students who were involved in bully/victim problems at school (Table VII). The prevalence was highest among boys in the age 14 group (grade 8), with 22.3% being involved either as a pure victim (10.2%), a pure bully (10.4%), or a bully-victim (1.7%). In grades 5 and 6, the percentages were around 15. For boys, but not for girls, involvement in bully/victim problems tended to increase with age. To provide a clearer overview, the main results are presented in bar graphs in Figure 3.

**DISCUSSION**

As explained in the “Introduction” section, there are a number of problems with several methods that have been, and are being, used for prevalence estimation in the area of victimization/bullying. It was argued that from the perspective of interpretability or meaning and reproducibility, use of a single (self-report) variable/item with quite specific response alternatives may be the “method of choice.” An important precondition for such a conclusion, however, is that the variable(s)/item(s) concerned can be shown to “function reasonably well” for their intended purposes. Such functionality includes indications of adequate construct validity and desirable psychometric properties. We start the “Discussion” section with a consideration of the functionality of the two key variables—the global questions about being bullied and bullying other students. After that, we turn to the issue of choosing a suitable cutoff point.

**The Functionality of the Two Key Variables**

In all analyses, we obtained the expected relationships between the global victim/bully variables and conceptually related variables. In all t-tests and ANOVAs, being bullied was markedly associated with the three “dependent” variables of internalizing problems—the social disintegration, the global negative self-evaluations (self-esteem), and the depressive tendencies scales. This was the case for girls and boys, separately and in combination. Similarly, bullying other students was strongly related to the two variables of externalizing problems, the general aggression and the antisocial behavior scales. The relationships in all these analyses were essentially linear, with higher levels or frequencies of being bullied/ bullying other students being associated with substantially higher levels in the dependent variables. Highly significant linear trends were found for girls and boys, separately and in combination, and for students in primary as well as in lower secondary grades.

All differences between the main groupings, between “victims” and “non-victims” and “bullies” and “non-bullies” (with “2-3 times a month” as a lower-bound cutoff point), were
highly significant for girls and boys, separately and in combination. Also, almost all of the differences between groups representing adjacent response categories (Group 0 vs. Group 1 or Group 1 vs. Group 2, etc.) were highly significant. Obviously, the results obtained were very robust and demonstrated a high degree of generality.

As is well known, statistical significance is to a considerable degree a function of sample size, and with a substantial sample as in the present study, effect size measures are more useful for judging the importance and magnitude of the differences obtained. It is common to
regard values of Cohen’s $d$ of .20, .50, and .80 as “small,” “medium,” and “large” differences, respectively, following Cohen’s cautious suggestion [see e.g., Cohen, 1988]. For the key comparisons in our study, the differences between “victims” and “non-victims” and between “bullies” and “non-bullies,” Cohen’s $d$ was very large for most of the dependent variables. For being bullied, the $d$ values varied between 1.05 for social disintegration and 0.62 for global negative self-evaluations, and for bullying other students, both $d$ values exceeded 1.0 for the aggression and antisocial behavior scales (boys and girls combined).

The very marked results in terms of $d$ indicate that the (point-biserial) correlations reported in the context of the linearity testing actually underestimate the strength of the relationships owing to skewness of distributions of the variables involved. To illustrate, with a $d$ value of 1.00 and an assumed 50/50 split of the victim/bully variables (representing the difference in, for example, social disintegration between equal-sized groups of victimized and non-victimized students), a point-serial correlation of around .30 (with a 10/90 split) would increase to approximately .45 [see e.g., Table 2.2.1 in Cohen, 1988].

Also, all of the cross-over correlations of the victim/bully variables with scales they were not expected to correlate with were close to zero and considerably smaller than the comparable correlation values for the conceptually related pairs of variables. These results attest to the discriminant validity of the key global variables and suggest that the obtained, expected relationships cannot, or only to a minimal degree, be explained as a function of shared method variance.

Finally, the school-level correlations between prevalence estimates derived from the dichotomized global measures and the dichotomized specific measures of victimization/bullying were quite substantial, approaching .80. This result shows that the school-level prevalence estimates were able to reliably differentiate schools from one another. This is of course a very desirable psychometric property of the two global variables. The considerable variation in prevalence rates among schools is a further indication of the same point.

Summing up, it can be concluded that all of the analyses conducted attested strongly to the functionality of the two key variables, both in terms of construct validity and psychometric properties.

Choosing a Suitable Cutoff Point

When using frequency of being bullied/bullying other students to identify possible victims and bullies, we need to decide how often a student must have been bullied/bullied other students to be classified as a victim or a bully. It is not self-evident how strict such a criterion should be. When making this decision, we emphasize both the results of the empirical analyses and conceptual considerations. In addition, we consider what may be called strategic arguments.

Is “2 or 3 times a month” a reasonable lower-bound cutoff point? As mentioned in the “Introduction” section, our point of departure was to set the cutoff point between the response categories “once and twice” and “2 or 3 times a month.” This was owing to conceptual considerations and previous traditions of both research and intervention/prevention. Our first task, then, is to assess to what extent this a priori cutoff point is a reasonable and useful one.

In the previous section, we emphasized that the differences between the main groups, between the “victims” and “non-victims” and “bullies” and “non-bullies,” reflected in the Group 2&3 vs Group 0&1 differences, were highly significant and very marked for all
psychosocial adjustment variables. The students classified as victims were clearly much more socially disintegrated (felt much less well liked by their peers, etc.), evaluated themselves much more negatively, and tended to be considerably more depressed than students who were not classified as victims (Table I). Similarly, students who were denoted “bullies” according to the chosen threshold value reported much higher scores on the general aggression and the antisocial behavior scales than students classified as “non-bullies” (Table IV). By and large, similar results were obtained for boys and girls, and the effect sizes indicated very substantial differences.

These analyses were followed up with comparisons between the two groups on either side of the cutoff point, Group 2 (“2 or 3 times a month”) vs. Group 1 (“only once or twice”). Also, these differences, which represent a stronger test of the chosen cutoff point, were highly significant for all dependent variables (Tables II and V). Because of smaller group sizes and the involvement of only directly adjacent response categories on the global victim/bully variables, these differences were, as expected, not as marked in terms of $P$ and $d$ values as in the previous comparisons.

All of these analyses as well as conceptual considerations strongly support our a priori, traditional choice of “2 or 3 times a month” as a lower-bound cutoff point for classifying students as victims and bullies. However, this conclusion should be regarded as tentative until we have examined the two alternative cutoff points of “only once or twice” and “about once a week.”

**What about the “only once or twice” category?** Also in the analyses comparing Group 1 (“only once and twice”) and Group 0 (“not bullied”), there were highly significant differences for all dependent variables. These results show that the psychosocial adjustment of students who admitted to having been bullied/bullied other students “only once or twice” were on average clearly different from that of students who had not been bullied/not bullied other students at all.

However, conceptual considerations may to some extent argue against use of this response category as a lower-bound value for classifying a student as a victim or a bully. The response category “only once or twice” does not agree well with the general emphasis on bullying as harassment that goes on over time and with some repetitiveness. The concerns about using this cutoff point are strengthened by the results from Table III, which suggest that the students in the category of “only once or twice” (Group 1) had on average been exposed to more temporary and less serious harassment than the students in the “2 or 3 times a month” category (Group 2).

In addition, nearly one third of the students, 31.2%, who replied “only once or twice” on the global question answered that they had not been bullied when asked about the duration of the bullying. Also, in the higher response category, “2 or 3 times a month,” there were some such apparently inconsistent results, 9.0%, but this figure is again considerably lower than the figure for the “only once or twice” group.

Such results may come about in several different ways. First, it may be an indication that the respondents felt somewhat insecure about what they had experienced: was it bullying or was it not? It is quite conceivable that in some cases this may be difficult to say. A second possibility is that not all students had fully understood the explanation of bullying included in the Questionnaire. Having a somewhat unclear feeling of what is meant by bullying may also lead to inconsistent responding. Third, because having been bullied may be experienced as a defeat or failure, it is also possible that some students admitted to being bullied on the global question but then wanted to “undo” their responding by negating the fact on one or more
follow-up questions. Finally, there is a possibility that some students were not quite serious when answering the questionnaire and hence their responding became erratic and inconsistent. In any case, it is obvious from Table III that a considerably smaller proportion of students who reported having been bullied “2 or 3 times a month” rather than “only once or twice” responded to the follow-up question in a way that was in apparent conflict with their earlier responses. In the higher response category, there was thus a larger percentage of students who seemed to “define themselves” as clear victims of bullying than in the lower category. This tendency was even more marked for the “about once a week or more” category, with only 3.6% giving the “not been bullied” response on the follow-up question about duration.

Summing up, in terms of psychosocial adjustment, the students in the “only once or twice” category (Group 1) were on average clearly different from the students in both the next higher (Group 2) and next lower (Group 0) groups. At the same time, the data on the follow-up question about the duration of the bullying suggested that a larger proportion of students in the “only once or twice” category had been exposed to more temporary and “lighter” or less serious harassment than students in the “2 or 3 times a month” category. All considered, it is natural to regard the response category having been bullied “only once or twice” as a measure of tendencies of being bullied rather than as a measure of clear victim status.

Similar results were obtained for the counterpart category, bullying other students “only once or twice.” However, for this variable, no parallel information about the duration of the bullying was available. But from Table V we see that this group was on average clearly different with regard to externalizing problems from both the students in the next lower (Group 0) and the next higher (Group 2) groups. Accordingly, it is also natural to use this response category for the global bullying variable as a measure of tendencies of bullying other students.

What about the “about once a week or more often” category? From Table II we see that the students who had been bullied “about once a week or more often,” Group 3, were significantly different from the Group 2 students (“2 or 3 times a month”) on all three internalizing variables. Similar results were obtained with regard to the two externalizing variables for corresponding groups on the bullying variable (Table V).

Regarding the follow-up question about how long the bullying had lasted, as many as 82.2% of the students from Group 3 (“about once a week” or more often) reported having been bullied “about a month” or longer. This percentage may be compared with corresponding values for “2 or 3 times a month” and “only once or twice,” that is, 64% and 24.9%, respectively.

These results might be taken as an indication that “about once a week” is a suitable lower-bound category for classifying students as victims and/or bullies. Use of such a cutoff point would certainly be in line with the general definition of bullying and also result in the selection of students with very clear victim and/or bully characteristics. However, such a choice would lead to several undesirable consequences.

First, it would fail to include in the (clear) victim or bully categories the relatively large groups of students who report being involved in bullying “2 or 3 times a month.” With “about once a week” as a lower-bound cutoff point, the “2 or 3 times a month” students would thus constitute two relatively numerous groups of “false negatives” in the framework of prediction terminology, that is, students who would be designated “non-victims”/“non-bullies” but actually have marked victim/bully characteristics.
Strategic considerations. Second, choice of a very strict criterion for classification of students as victims or bullies would likely reinforce a view of victimization/bullying as something that only happens to “extreme” and presumably highly disturbed students. This may result in “victim blaming” and, more generally, in an excessive focus on individual students rather than on contextual factors and the school as a social environment that can facilitate or counteract bullying. Third, it would make it more difficult to detect tendencies toward bullying at an early stage, which is quite important from a prevention perspective. In a similar vein, schools with such a view would probably be less willing to introduce anti-bullying measures or programs that are directed toward all students in the school. As documented elsewhere, anti-bullying programs of this “universal” kind have proven to be effective [see e.g., Olweus, 1991, 1994; Olweus and Limber, 1999; Smith and Sharp, 1994], and such programs are more likely to be accepted by schools than programs that aim at particular high-risk groups [Olweus, 1992].

On the other hand, the classification criterion should not be too inclusive either. If we chose “only once or twice” as our lower-bound category, we would have problems with relatively large groups of “false positives,” that is, students who would be classified as victims or bullies but who did not have very marked victim or bully characteristics. In addition, we would have some difficulty reconciling the obtained data on the length (and repetitiveness) of the bullying with our general definition of bullying, as explained previously. In addition, high prevalence rates may have the unfortunate effect of “normalizing” or “trivializing” bullying, thereby reducing efforts to combat bullying: bullying is seen as a natural and normal part of growing up, as something every child meets and must learn to handle on their own.

Conclusion about cutoff point. On the basis of our empirical analyses, conceptual arguments, and strategic considerations, we arrive at the conclusion that the response category “2 or 3 times a month” (in the past couple of months) is a reasonable and useful lower-bound cutoff point for classifying girls and boys in the 10 to 16 years age range as victims and/or bullies for purposes of prevalence estimation. As mentioned in the “Overview of Article” section, this was our point of departure and a priori choice. Nothing in our reported analyses has contradicted this choice of cutoff point, and a good deal of new evidence has been provided in favor of it. Accordingly, we recommend that the percentages of students having been bullied and/or bullied other students “2 or 3 times a month” or more often should be regularly reported and be the basic prevalence measure in studies using the Olweus Bully/Victim Questionnaire for these and adjacent age groups. If this practice is followed, a useful standard of reference will be established that will facilitate comparisons among units such as schools, communities, or even countries at a particular point in time and across time.

At the same time, it may be of value also to report the percentages of students who have been bullied or bullied other students “about once a week” or more often and/or “only once or twice,” at least at a summary level (cf. Table VI). Such prevalence estimates will provide useful information about the frequency of more serious cases, on the one hand, and the frequency of lighter cases, on the other, representing tendencies of being bullied and bullying other students, respectively. Such additional information may be helpful when schools or larger units deliberate or make plans for prevention/intervention.

Although these guidelines are straightforward and clearly meaningful in the context of the analyses presented, it should be cautioned that prevalence figures from different countries or different cultures may not be directly comparable. Despite the fact that the Questionnaire gives a detailed definition of bullying, it is likely that the prevalence rates obtained will to
"some extent be affected by translation difficulties (whether or not there is an appropriate word for bullying in the language), the students’ familiarity with the concept of bullying, the degree of public attention to the phenomenon, and similar factors. Accordingly, before drawing conclusions about prevalence estimates from different countries/cultures, a number of “psychometric” analyses such as those presented herein should be undertaken. This is to assess the degree to which the meaning and implications of “being bullied” and “bullying other students” are similar and comparable across the contexts of interest.

Brief Comments on the Empirical Prevalence Results

As mentioned in the “Introduction” section, the empirically derived prevalence rates are not in themselves a major focus in this context but rather serve as an illustration of the more general and basic issues dealt with in the article. The results obtained are certainly interesting in themselves, however, and readers who want more detail are referred to Tables VI and VII, Figure 3, and the associated text in the “Results” section.

Here we just want to stress the value of reporting prevalence data not only for victims and bullies as overall categories but also for the more circumscribed groups of victims only (pure victims), bully-victims, and bullies only (pure bullies); see also Wolke et al. [2001]. Because bully-victims, or provocative/aggressive victims, which they have also been called, are a group with special characteristics [e.g., Austin and Joseph, 1996; Craig, 1998; Haynie et al., 2001; Kumpulainen et al., 1998; Nansel et al., 2001; Olweus, 1973, 1978, 1993; O’Moore and Kirkham, 2001; Schwartz et al., 2001], separating them from the overall groups of victims and bullies in statistical analyses will often give stronger and more easily interpretable results.

Concluding Comments

The main objectives of this article have been to study the functionality of the two global variables in the Olweus Bully/Victim Questionnaire and the appropriateness of a special cutoff point on these variables for purposes of prevalence estimation. A number of empirical and conceptual analyses have strongly attested to the functionality of the two variables and, in combination with strategic arguments, have indicated that “2 or 3 times a month” is a reasonable and useful lower-bound cutoff point for purposes of prevalence estimation. Prevalence estimates derived in this way can be conveniently obtained, have a reasonably well-defined meaning, can be easily understood by the users, and can be reproduced unambiguously by different researchers/administrators and in different time periods (maybe within certain limits).

Although these aspects certainly represent important advantages, it is also natural to ask whether there are particular problems associated with use of this self-report strategy. There are two potential problems that may be of special relevance in this context. First, one must consider the possibility that the construct validation results (shown, e.g., in Figs. 1 and 2) were partly or wholly a consequence of shared method variance, since only one data source was used. Our analyses, however, suggest that this may not be a big problem since the cross-over correlations were close to zero and the two (dichotomized) key variables correlated only .13 with each other (girls and boys combined).

Another problem is related to possible underreporting or overreporting. This is a big and complex issue that cannot be treated satisfactorily within the scope of the present article [see, e.g., Björkqvist et al., 1992; Österman et al., 1994; Perry et al., 1988; Salmivalli et al., 1996b; Schuster, 1999; Smith and Levan, 1995; Smith et al., 1999a; Sutton and Smith, 1999].
However, the most obvious possibility is that some children who bully other students will choose not to respond honestly on the relevant Questionnaire items. Although this cannot be ruled out, we want to remind the reader that the relationship between the bullying variable and the two externalizing variables was (at least) as strong as the relationship between the victimization variable and its internalizing correlates, thereby contradicting the hypothesis that social desirability exerts a major influence in the case of the bullying variable. Of relevance in this context and a possible partial explanation is the fact that the students responded anonymously to the Questionnaire [see, e.g., Ahmad and Smith, 1994] and that the confidentiality of the responses was (and is) strongly emphasized in the instructions for administration. Nonetheless, more research on this issue is needed before definitive conclusions can be drawn.

As has been stressed several times in this article, the attractiveness of single items/variables is closely linked to prevalence estimation for a larger unit such as a school or a community. For other purposes, as in attempts to “validate” bully/victim information from different sources of data against each other, single-item variables are often insufficient. In such contexts, it is important to employ as reliable information as possible, e.g., carefully constructed indices or scales built on a number of items. With the present Questionnaire, this can be achieved by creating a sum or preferably an average of the responses to the global question and the following seven (or eight) questions about various forms of victimization/bullying (and maybe additional items). Such indices or scales are also to be preferred, e.g., when the goal is to identify groups of victims and bullies (and non-involved students and maybe bully-victims) with marked victim and/or bully characteristics. In such studies, the usefulness and consequences of possibly adding and integrating information from other data sources such as peer or maybe teacher nominations or ratings should also be explored in more detail than has been done so far.

ACKNOWLEDGMENT

This research was conducted as part of Mona E. Solberg’s fulfillment of the requirements for the Doctoral Degree of Philosophy at the University of Bergen, Norway, under the supervision of Dan Olweus. The research program reported on in this article was supported by grants from the Norwegian Research Council (NFR) to Mona E. Solberg and Dan Olweus and grants to Dan Olweus from the Norwegian Foundation for Health and Rehabilitation/the National Association for Public Health, the Ministry for Children and Family Affairs (BFD), the Johann Jacobs Foundation, Switzerland, and, in earlier phases, the Norwegian Ministry of Education (KUD).

REFERENCES


