Abstract: Currently, the UIS reports estimates of SDG 4.a.2 using data from two school based surveys (GSHS and HBSC). This note builds on methodology proposed by UIS and UNICEF in 2018 by proposing an estimation method that uses student background data from international student assessments, in addition to GSHS and HBSC data. Because of limitations in the comparability of estimates across the different survey programmes, the proposed approach for reporting a single monitoring indicator is to report an indicator in five-year bounds by, first, ranking the survey programmes based on how closely their estimates match SDG 4.a.2 and, second, using estimates for each country and sub-population from the highest ranked survey programme. By using only one survey programme as the source of data per country, this ensures cross-time comparability to monitor each country’s progress which is the primary goal of the SDG indicator, rather than for making cross country comparisons. It is also proposes that UIS share estimates of the indicator for each country and survey programme for use by researchers.

Background

Research has documented the substantial adverse effects of school-age bullying on health and educational outcomes. Bullying during school years adversely affects physical and mental health (Moore et al. 2017; Wolke & Lereya 2015) through increased risk of depression, anxiety, loneliness and sadness as well as suicide (Moore et al. 2018; Kochel, Ladd & Rudolph, 2012; Livingston et al. 2019; Rigby and Cox 1996). Using PISA 2018 data, OECD (2019) found that students who were frequently bullied were more likely to report feeling sad, scared and not satisfied with their lives. Exposure to bullying has been linked to lower educational outcomes
Brendgen 2018; Espelage & Colbert 2016) for both the victims and aggressors of bullying through an increased risk of missing classes, dropping out of school and lower cognitive achievement (Juvonen, Yueyan Wang & Espinoza 2011; Konishi et al. 2010; Townsend et al., 2008). The longer term education and economic consequences have resulted in bullying becoming an important policy issue for governments internationally (Morrow, Barnett, & Vujcich 2013; Peyton, Ransingh & Jacobsen 2017; Phillips 2007).

**Bullying in school is generally defined to include verbal and relational abuse in addition to physical abuse.** It exploits an unequal power relationship in which the bully harms or discomforts a victim (Olweus 1993; Woods & Wolke 2004). Bullying can be characterized into physical abuse, verbal abuse or relational which includes spreading rumours, public humiliation, shaming and social exclusion (Woods & Wolke 2004 in OECD 2019). Cyber-bullying has become another common form of bullying in which abuse, particularly relational abuse, utilizes technology including spreading rumours online or excluding someone from an online group (Hinduja & Patchin 2010; Smith et al., 2008; OECD 2017). It differs from previous forms of bullying because of bully’s ability to remain anonymous in many cases and the fact that bullying can take place anywhere or anytime depriving the victim of recourse or the ability to escape (Slonje & Smith 2008; Wang, Iannotti & Nansel, 2009).

The prevalence of school bullying varies across counties but all children in all countries are at risk; little research has studied low and middle income country contexts. The prevalence of bullying depends on a range of factors including the child’s age, country and cultural factors, relational factors, environmental factors and the type of study the child is enrolled in pursued (Chester et al. 2015; Craig et al. 2009; Saarento, Garandeau & Salmivalli 2015); however, all children in all countries can be affected (Nansel et al. 2004). While the prevalence, causes and effects of school bullying has been studied in high income country contexts, there are few studies about school bullying in low and middle income countries (Biswas et al. 2020).

Currently, the UIS reports an indicator of SDG 4.a.2 based on two school health surveys, the GSHS and HBSC. Potential data sources for an SDG 4.a.2 indicator were assessed by the UIS (2018). This review studied the possibility of using the Health Behavior in School-Aged
Children (HBSC) survey and the WHO’s Global School-based Student Health Survey (GSHS) as well as international student assessment surveys (Table 1). The HBSC collects data on children aged 11, 13 and 15 from Europe and North America, while the GSHS collects data on children aged 13 to 17 for more than 80 countries, globally. Both surveys collect data on bullying including verbal, physical and relational abuse. In their questionnaires, they define bullying and ask how often the respondent (student) has experienced any type of bullying in the past couple of months for the HBSC and past 30 days for the GSHS. The HBSC 2013-14 round also asks whether the respondent has experienced cyber-bullying; the GSHS, for all rounds since 2003 asks which types of bullying, including physical bullying, being made fun of, and being left-out, were most often experienced. The UIS reports the percent of students who were bullied based on these two surveys. Disaggregation by sex, immigrant background and socio-economic status are also reported. The UIS study extends work by UNICEF Innocenti (Richardson & Hiu 2018) who have proposed a methodology for developing a global indicator using a combination of the HBSC and GSHS surveys as well as CNAs TIMSS and LLECE.

<table>
<thead>
<tr>
<th>Survey</th>
<th>Target population</th>
<th>Scope of bullying</th>
<th>Asks about bullying in the past...</th>
<th>Years</th>
<th>Number of countries</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSHS</td>
<td>students aged 13-17</td>
<td>Saying or doing bad and unpleasant things, teased a lot in an unpleasant way or when a student is left out of things on purpose. Prompted specifically for types of physical bullying, being made fun of, and being left out</td>
<td>Past 30 days</td>
<td>Rounds: 2003-2008, 2009-2012, 2013-present</td>
<td>101 countries in at least one round (global coverage)</td>
</tr>
<tr>
<td>HBSC</td>
<td>11, 13 and 15 year-old students</td>
<td>As GSHS, except only specific prompts for cyber-bullying</td>
<td>Past 2 months</td>
<td>Latest round: 2013/14*</td>
<td>42 countries in Europe, North America (and Israel)</td>
</tr>
</tbody>
</table>

*Only the HBSC 2013/14 questionnaires were available for review for this study.

As documented in UIS (2018), latest rounds of the international student assessments, TIMSS, PISA, and, LLECE, collect data on exposure to physical, verbal and relational
bullying. The scope of bullying included in the latest rounds of TIMSS, PISA and LLECE were similar, reflecting physical, verbal and relational forms of bullying (Table 2). TIMSS emphasized the spreading of embarrassing information as well as including online abuse as a separate question, while PISA explicitly mentions online forms of harassment to be considered when responding to the set of questions on bullying. LLECE does not explicitly mention online forms of bullying but also does not exclude them. PASEC’s student questionnaire asks whether student feel safe or scared at school as well as whether they experience violence or corporal punishment; hence, this question about safety may capture violence from other students but also from teachers.

<table>
<thead>
<tr>
<th>Survey</th>
<th>Sampled population</th>
<th>Scope of bullying questions</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>LLECE 2013 (TERCE)</td>
<td>6th grade</td>
<td>afraid of, felt threatened by, fear of violence from, made fun of by, excluded by, forced to do things by classmates</td>
<td>ever happened while you were at school</td>
</tr>
<tr>
<td>PASEC 2014</td>
<td>6th grade</td>
<td>felt scared, not safe in class</td>
<td>in general at school</td>
</tr>
<tr>
<td>PISA 2018</td>
<td>15 years-old</td>
<td>excluded, made fun of, threatened, property stolen or damaged, hit or pushed around, spread nasty rumours</td>
<td>a few times a year or month, at least once a week</td>
</tr>
<tr>
<td>TIMSS 2015</td>
<td>8th grade</td>
<td>made fun of, spread lies, stole something, hit or hurt me, made to do things, shared embarrassing information, posted embarrassing things, threatened me</td>
<td>a few times a year, at least once per month or week</td>
</tr>
<tr>
<td>TIMSS 2015</td>
<td>4th grade</td>
<td>as for 8th grade</td>
<td>a few times a year, at least once per month or week</td>
</tr>
</tbody>
</table>

See meta-data (Annexe 1) for details on questions in each survey.

Main methodological decisions

1. Advantages of adding CNA data and comparability with GSHS / HBSC: Measurement of SDG 4.a.2 using LLECE, PISA, and TIMSS offers some advantages to complement the current
indicator based on the GSHS and HBSC. First, PISA and TIMSS collect data on bullying that has happened in the past year which reflects the timeframe of SDG 4.a.2; the GSHS and HBSC ask about bullying that has occurred only in the past month or two. Second, TIMSS offer an earlier point of measurement at 4th grade, and LLECE, TIMSS and PISA allow for estimating the prevalence of bullying by primary or secondary school level which is helpful for policy makers designing and targeting programmes. GSHS does not explicitly mention cyber-bullying, while PISA and TIMSS do; this allows for the learning assessments to provide measurement on a wider scope of bullying in countries not included in the HBSC (which is North America, Europe, and Israel).

2. Frequency of bullying: HBSC and GSHS ask about bullying in the past month or two. PISA and TIMSS ask about how often the student respondent has experienced bullying ranging from once in the past year once or more a week. It is possible to develop an indicator using the PISA and TIMSS data on whether the student has experienced bullying in the past month in order to match the GSHS and more closely match the HBSC. However, the SDG indicator explicitly states bullying in the past 12 months; hence, the proposed indicator derived from student assessment data would measure bullying in the past 12 months while the indicator derived from GSHS would denote bullying in the past 30 days and, for the HBSC, the past two months.

3. Scope of bullying: Research typically describes bullying to include physical, verbal and relational aspects. As a result, the full scope of bullying defined in the surveys would be included in the definition of the indicator. This may limit some comparability between surveys, particularly for LLECE 2013 and GSHS which do not mention or prompt for cyberbullying. An alternative would be to report only the forms of bullying that are common to all surveys; however, it is not possible to exclude cyberbullying in the PISA or HBSC data. Note that the student assessment questionnaires require students to answer whether they have been subjected to different types of bullying separately. In the HBSC and GSHS core questionnaires, bullying is defined similarly but students are asked how often they have experience any type of bullying. GSHS asked subsequently which type of bullying they were exposed to most, while HBSC (2013-14) asks about whether students have been subjected to two forms of cyber-bullying. While the survey programmes define bullying similarly, differences in asking about specific
forms of bullying may prompt students to respond differently. For example, if a student was exposed to relational bullying, he or she may not answer yes unless specifically asked whether he or she was intentionally left out of activities.

4. Measurement points and baseline definition: Reporting bullying from the CNAs and HBSC/GSHS data would offer multiple points of measure for different target populations. One approach may be to combine these data in order to provide one indicator per country; however, this would result in an indicator that varies by sub-population, scope of bullying and frequency of bullying across countries and time depending on what survey data is available for a country. This would also not be useful for researchers as the source data would not be easily available from UIS. Because survey data is being used to estimate this indicator, the resulting values are actually estimates of the SDG 4.a.2 indicator rather than official values. ILOSTAT relies heavily on estimates for employment statistics and, as a result, reports indicators by country and survey. This approach is proposed for UIS’s reported of estimates for SDG 4.a.2 from the CNA data and HBSC/GSHS data. For the purposes of monitoring, the proposed approach is to determine a ranking of survey programmes and to report the value of the highest ranked survey that a country participates in. Because the primary purpose of the SDG indicator is to monitor a country’s progress across time rather than to compare between countries, this approach of using one survey programme per country would best ensure comparability across time. The protocol for reporting is described in detail below.

Proposed indicator

Purpose: The proposed indicator aims to define bullying as including physical, verbal and relational abuse. This scope reflects current research on bullying as well as the definitions for major international student assessments.

Definition: The indicator would be defined as the percent of students who have experienced any type of bullying in the past year, by sex, following UIS (2018) and Richardson & Hiu (2018). For assessment $i$, the measure of prevalence of bullying for the assessment’s target population $\beta_i$
would be defined as
\[ B_i = E[1\{b_1 = 1 \text{ or } b_2 = 1 \text{ or } ... \text{ or } b_{n_i} = 1\}] \] (1.1)

where \(1\{\ldots\}\) denotes the indicator function which takes value 1 if the expression is true and zero if false. Variables \(b_1, \ldots, b_{n_i}\) denote the various types of bullying included in the question on what types of bullying the student may have experienced; these variables equal 1 if the student has experienced the type of bullying and zero if the student answers no. \(E[\ldots]\) denotes the population mean (expected value); the methodology for estimating the expected value of \(1\{b_1 = 1 \text{ or } b_2 = 1 \text{ or } ... \text{ or } b_{n_i} = 1\}\) varies by survey and depends on the survey’s sampling design. Students who did not answer any of the bullying questions would be omitted from the calculation; for students that omitted some of the questions, the omissions would be treated as zeros. The target population would be that of the assessment but excluding those unwilling to answer any of the bullying questions. The time period, to match the SDG indicator definition, would be whether bullying was experienced at least once in a year.

**Summary of estimated indicators using assessment data globally**

**Prevalence of bullying estimated in the assessment data ranges from 20 to 95 percent depending in the measurement point, country and source of data.** The proposed indicators were estimated for LLECE 2013, TIMSS 2015, and PISA 2018 as described above (Figure 1). The range in the percent of students being exposed to bullying varies considerably by country. For children (LLECE 2013 3rd grade and TIMSS 2015 4th grade), the prevalence of bullying ranges from 30 to 88 percent; at the adolescent level (PISA 2018 and TIMSS 2015 8th grade), the range is from 20 to 95 percent. The prevalence of bullying estimated using TIMSS data tends to be higher than that estimates using the other sources, especially at the child level. Comparing countries that were included in both TIMSS 2015 8th grade and PISA 2018, large differences in the percentages are found (Figure 2). Of the 29 countries that were included in both surveys, the differences range from -1 to 42 percentage points, with an average of a 20 percentage point difference. At the child level only two countries with bullying prevalence estimates were sampled both in LLECE 2013 and TIMSS 2014. The differences in TIMSS and the other data
sources may be explained by differences in the question item on bullying. TIMSS included an explicit question on cyber-bullying, and it did not restrict the place of bullying to the school. Another difference is age. Children in 8th grade may be younger than 15 which may result in higher prevalence of bullying. Note that the most common forms of bullying found in TIMSS were spreading lies and being made fun of; for the whole sample of all countries, 36 percent of the sampled students (un-weighted) reported not being affected by either of these two. The OECD (2019) reports the percent of students being bullied a few times a month rather than in the last year, which provides as a result much lower levels of prevalence.
Fig. 1. Percent of students exposed to bullying by country

Source: author’s calculations using LLECE, PISA and TIMSS data.
On average, the prevalence of bullying among boys is 4.9 percentage points higher than girls with differences ranging from -5.6 to 18.8 percentage points by country. Higher prevalence of bullying among boys was found for all survey programmes (Figure 3). Of the 196 datasets, boys reported a higher prevalence of bullying in 170 while girls reported a higher prevalence in 26 datasets. At the primary level, the average gender difference (between boys’ and girls’ bullying prevalence) across countries was 1.0 and 4.5 percentage points for LLECE 2013 6th grade and TIMSS 2015 4th grade, respectively. At the secondary level, the average gender difference across countries was 6.7 and 3.8 percent for PISA 2018 and TIMSS 2015 8th grade, respectively. No systematic large differences in gender difference were found between estimates of the gender difference in bullying across datasets were found.
Fig. 3. Percentage point gender difference in exposure to bullying by country (male exposure – female exposure)

Source: author’s calculations using LLECE, PISA and TIMSS data.
Comparability of estimates across surveys

Reliability and comparability of assessment data estimates: As with any measure using multiple CNAs, comparability is limited by differences in the surveys’ target populations. For the proposed indicator, comparability is also limited by (1) the additional question in TIMSS on whether the child has experienced the sharing embarrassing information, (2) the explicit question on on-line bullying in TIMSS, (3) the lack of time period specified in LLECE, (4) varying ages within grades which affect the prevalence of bullying, and (5) place of bullying. The additional question in TIMSS on whether the child has experienced the sharing of embarrassing information may be implicit in the PISA questionnaire, but it may capture a slightly wider scope of bullying resulting in a higher prevalence of bullying compared to PISA. The explicit question on on-line bullying in TIMSS may also, by prompting, result in a more reliable measure of cyber-bullying compared to PISA where it is explicitly mentioned in applying to questions on all forms of bullying or compared to LLECE (TERCE) which does not mention on-line bullying. Both PISA and TIMSS ask about the child’s exposure to bullying as whether it occurred in the past year, past month or past week. LLECE (TERCE) does not specify a time period; as a result, the TERCE item may include bullying that occurred more than a year ago. While PISA samples only 15 year-olds, TIMSS and LLECE sample by grade which results in variation in ages; countries may differ on the prevalence of varying due to the age profile within grades. PISA and LLECE 2013 ask whether bullying has taken place at school while TIMSS does not specify the location of bullying, but rather, asks about whether bullying was perpetrated by students from the school. Finally, the answer for no bullying in PISA is “never or almost never” which may inflated the percent of students that have not been bullied compared to TIMSS.

Differences in data collection, target populations, and timeframe for bullying yields significant differences in estimates of bullying between survey programmes; this limits the comparability of estimates between survey programmes. There is significant variation in bullying across countries measured in the three assessment programmes (Figure 1 & 3). Analysis of the HSBC survey found much lower prevalence of bullying with only a few of the 33 countries or regions reporting rates higher than 50 percent (Chester, K. et al. 2015). This is likely the result of the reporting period; HSBC and GSHS ask about bullying in the past month
or two. There are also important differences between the definitions of bullying in the learning assessment data compared to the HSBC survey as discussed above. While the CNAs offer to increase the number of data points and the breadth of bullying, and allow in the case of TIMSS and PISA to ask about bullying in the past year, their comparability both with each other and with the HSBC and GSHS (which asks about bullying in the past month) is limited.

Protocol for reporting the indicator

**Measurement points:** Two measurement points are proposed: (1) primary level that includes assessments targeting populations lower than 8th grade or younger than age 13, (2) secondary level that includes assessments targeting populations 8th grade and above or age 13 and higher (Table 3). This is an approximate classification because in some cases 8th grade may belong to primary school.

<table>
<thead>
<tr>
<th>Measurement point</th>
<th>Assessment ranking</th>
<th>Rationale for ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary</td>
<td>TIMSS 2015 4th grade</td>
<td>Specified bullying in the past year; includes cyber-bullying explicitly</td>
</tr>
<tr>
<td></td>
<td>HBSC age 11</td>
<td>includes cyber-bullying explicitly</td>
</tr>
<tr>
<td></td>
<td>LLECE 2013 6th grade</td>
<td>n.a.</td>
</tr>
<tr>
<td>Secondary</td>
<td>TIMSS 2015 8th grade</td>
<td>Specified bullying in the past year; includes cyber-bullying explicitly. Has higher estimates of bullying than PISA suggesting wider scope in practice.</td>
</tr>
<tr>
<td></td>
<td>PISA 2018</td>
<td>Specified bullying in the past year; includes cyber-bullying explicitly.</td>
</tr>
<tr>
<td></td>
<td>HBSC age 13 and 15</td>
<td>Includes cyber-bullying explicitly</td>
</tr>
<tr>
<td></td>
<td>GSHS</td>
<td>n.a.</td>
</tr>
</tbody>
</table>

This table presents the rankings only for the assessments and surveys reviewed for this study; as more surveys are reviewed, this list would be updated. Note that LLECE 2013 3rd Grade does not collect bullying data; PASEC 2014 does not collect bullying data relevant to the scope of the proposed indicator.

**Reporting multiple estimates of the indicator:** Because comparability of estimates of bullying between survey programmes is limited, the proposed approach is to report, for each country and sub-region, estimates of the indicator from each survey programme separately. For example, the data would be structured by country, then sub-region, then year, then survey programme, then
target population, and finally for each of these, the estimated value. For example, there would be several estimates for China. One set of estimates would be reported for Hong Kong, China. For Hong Kong, China, in 2015, there would be two estimates for bullying: from 8th Grade TIMSS and PISA. Reporting multiple surveys per country follows the approach by ILOSTAT which relies heavily on estimated data for labour market outcomes from varying sources rather than official numbers.

**Monitoring indicators:** For the purposes of monitoring, a single estimate per country would be provided in five year bounds: 2011-2015, 2016-2020, 2021-2025, and 2026-2030. Because of limitations in the comparability of the bullying indicator across countries, indicator values for each country would use estimates only from the survey programme in which they participate that provides estimates that most closely match SDG 4.a.2. The ranking of surveys generally within each measurement point places TIMSS first because it tends to have highest estimates of bullying suggesting that it captures a wider scope than PISA in practice (Table 4). TIMSS and PISA rank higher than HBSC and GSHS because they report bullying within the past year, more closely matching the SDG indicator. HBSC is preferred to GSHS because of the explicit mention of cyber-bullying. Note that these rankings are based on the questions that are included in the core questionnaires which are assumed to be the most readily available. The advantage of using only a single assessment programme for a given country is that it requires not modeling methodology which would reduce the transparency of the indicator. A limitation is that countries may choose to stop participating in an assessment program in the future which would require changing the source of data for the indicator to preserve comparability. The only calculation that would need to occur with the data is when more than one estimate of the highest ranked assessment programme is available within a five-year time-bound; in which case, an average of estimates is proposed.

**Comparing with UNICEF’s proposal (Richardson & Hiu 2018)** -- The proposed indicator and methodology differs from that of Richardson and Hiu (2018) primarily in the reporting protocol. In order to account for differences in comparability across survey programmes, they use a modeling method to “normalize” estimates. This improves the comparability of estimates across countries; however, the limitation of using a modeling approach is that it reduces transparency
by making it difficult to explain to policy makers and the public. Because the purpose of the SDG monitoring indicator is to track each country’s progress rather than to compare across countries, using a single survey programme to report indicators best ensures cross-time comparability.

<table>
<thead>
<tr>
<th>Time bound</th>
<th>PISA</th>
<th>TIMSS</th>
<th>LLECE</th>
<th>GSHS</th>
<th>HSBS</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011-2015</td>
<td>avg. of 2012 &amp; 2015</td>
<td>avg. of 2011 &amp; 2015</td>
<td>only 2013</td>
<td>as available, depends on country participation</td>
<td></td>
</tr>
<tr>
<td>2016-2020</td>
<td>only 2018</td>
<td>only 2019</td>
<td>only 2019</td>
<td>TBD</td>
<td></td>
</tr>
<tr>
<td>2021-2025</td>
<td>avg. of 2021 &amp; 2024</td>
<td>only 2023</td>
<td>TBD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2026-2030</td>
<td>avg. of 2027 &amp; 2030</td>
<td>only 2027</td>
<td>TBD</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note this is assuming that data is available for a bullying indicator in the specified rounds*
References


Olweus, D. (1993), Bullying at School: What We Know and What We Can Do, Blackwell


UIS 2018. *SDG thematic indicator 4.a.2 – Safe, non-violent, inclusive and effective learning environments* Percentage of students who experienced bullying during the past 12 months, by sex. Montreal: UIS

Wolke, D. and S. Lereya (2015), Long-term effects of bullying, Archives of Disease in Childhood, Vol. 100/9, pp. 879-885