

Defining Bullying: A Split-Ballot Survey Experiment across Three Federal Departments

Melissa A. Cidade, Ph.D.¹

Avar Consulting, Inc.

Rachel Hansen¹

National Center for Education Statistics, Department of Education

Proceedings of the 2018 Federal Committee on Statistical Methodology (FCSM) Research Conference

Abstract

Bullying in America's schools is a primary concern for several federal agencies, including the Department of Education, the Department of Justice, and the Department of Health and Human Services, among others. However, examining different sources of bullying victimization data leads to different understandings of the magnitude of the issue. What these estimates obfuscate is that each of these surveys uses a different definition of bullying. To ameliorate the issue of apples-to-oranges comparisons in the rate of bullying victimization, a working group of federal partners collaborated to determine a uniform definition of bullying, published by the Center for Disease Control (CDC) in 2014. In response to this uniform definition, researchers at the Department of Education changed the way they asked about bullying in school on the 2015 collection of the School Crime Supplement (SCS) to the National Crime Victimization Survey (NCVS). Since the SCS has been conducted biennially since 1999, researchers had a quagmire: how do they continue to collect the trend data on national bullying estimates while bringing the questionnaire into alignment with the newly published CDC definition? The answer was to embed a split-ballot experiment on the 2015 administration of the SCS. Our project is to outline the methodological and institutional challenges to changing the way that bullying victimization is asked on the SCS. First, we will discuss the concepts embedded in the CDC definition, and ways to operationalize them to survey questions. Then, we will outline the research procedures used to develop and field the split-ballot experiment. We will conclude with results of the experiment, as well as discuss our experiences with best practices for coordinating changes in survey instruments across the three federal agencies involved in the administration of the SCS.

Introduction

Since its introduction to the academic and applied literature in the 1970s, the phenomenon of peer-to-peer student victimization – bullying – has become a centerpiece of educational, public health, and other social programming. Bullying victimization has been linked to negative social outcomes, and programmatic interventions have sought to reduce the behavior in elementary, middle, and secondary schools. However, a larger issue continues to plague bullying research and interventions: there is no standardized definition of bullying, and no set way of measuring the phenomenon in social research.

In response to the lack of uniformity in bullying measurement, in 2014, the Centers for Disease Control and Prevention (CDC) published a set of guidelines for recognizing and measuring bullying victimization. This definition, which brings together the concerns of the education, public health, and social services communities, emphasizes aggressive peer-to-peer victimization, power imbalance, and repetition. As a result of these new guidelines, one federal data collection effort – the School Crime Supplement (SCS) to the National Crime Victimization Survey (NCVS) – was out of alignment with the components of the CDC Uniform Definition. Researchers at three federal agencies, including the Department of Justice's (DOJ) Bureau of Justice Statistics (BJS), the Department of Education's (ED) National Center for Education Statistics (NCES), and the Department of Commerce's Census Bureau devised a randomized split-ballot survey experiment to test a newly operationalized bullying victimization question. The implementation of this experiment brought together those in bullying research with education, health, and social services practitioners. While the experiment was ultimately not continued into subsequent administrations of the survey, it did provide important insight into the coordination of several federal agencies focused on bullying prevention.

¹ This paper is intended to promote ideas. The views expressed are part of ongoing research and do not necessarily reflect the position of the U.S. Department of Education or Avar Consulting, Inc.

Strategic Importance of Bullying Prevention

In 2011, a Conference on Bullying Prevention was held at the White House where President Obama noted that “A third of middle school and high school students have reported being bullied during the school year. Almost 3 million students have said they were pushed, shoved, tripped, even spit on” (Lee 2011). This conference elevated the issue of bullying to a national conversation. That same week, bipartisan anti-bullying legislation was introduced in the Senate with the stated purpose of addressing “the national crisis of bullying and harassment” (S.506 2011). This piece of legislation has been reintroduced in 2013, 2015, and most recently in the House in 2017.

Bullying has been elevated as a critical area of investigation by several key federal agencies. Some explicitly state their attention to bullying. For example, ED states in Goal 4 of the 2014-2018 Strategic Plan a focus to “increase educational opportunities for underserved students and reduce discrimination,” along with an end to “discrimination, harassment, bullying, and other barriers to education” (U.S. Department of Education 2014: 33). Likewise, DOJ explicitly states in objective 2.5 of the 2014-2018 Strategic Plan a goal to “address education discrimination and segregation,” and more specifically, to ensure “that schools respond appropriately to harassment of students on the basis of sex, race, national origin, disability, and religion” (37). Others implicitly have turned their attention to peer victimization. For example, the Department of Health and Human Services (HHS) has prioritized “promot[ing] the safety, well-being, resilience, and healthy development of children and youth,” with a particular emphasis on programs that “attend to children’s behavioral, social, and emotional functioning” (2014).

To effectively use resources and to maximize the impact of anti-bullying research and programs, representatives from nine federal agencies formed the Federal Partners in Bullying Prevention Steering Committee. This partnership produced a clearinghouse of anti-bullying information and programming (hosted at www.stopbullying.gov), as well as hosting several anti-bullying summits. Other interagency efforts include the Safe Schools/Health Students Initiative (funded jointly by ED and HHS’s Substance Abuse and Mental Health Services Administration (SAMHSA)) which includes grants for local education agencies to implement activities, programs, and services focused on health childhood development and violence prevention, and efforts by federal agencies to publish guidelines on high-quality emergency operations plans, developed specifically for elementary and secondary schools (McCallion and Feder 2013: 11-12).

Defining Bullying

While the academic and policy literature on bullying is rife with studies on the impacts, predictors, and prevention of bullying, the fundamental question, “What *is* bullying?” remains unclear. Researchers and policy makers have, for the most part, defined the concept on their own as their projects demand. The lack of uniformity in the characteristics of bullying makes comparing studies on the topic inconsistent at best, and impossible at worst.

Swedish psychologist Daniel Olweus is largely credited with the introduction of the term “bullying” into the academic lexicon. In 1978, his book, *Aggression in the Schools: Bullies and Whipping Boys*, provided the framework for peer harassment and bullying, and is still the seminal work on the topic (Olweus 1978). Olweus later admitted that the need for “a relatively clear and circumscribed definition became urgent in connection with the government-initiated campaign against bullying in Norway in 1983” (Olweus 2013: 755). It is within the context of this campaign that the first codified definition of bullying was established. At that point, the Olweus definition of bullying was (and remains):

“A student is being bullied or victimized when he or she is exposed, repeatedly and over time, to negative actions on the part of one or more other students”
(Olweus 1993: 9).

Olweus further defines important terms in his definition of bullying. Negative actions include “when someone intentionally inflicts, or attempts to inflict, injury or discomfort upon another” (9); this inclusion introduces the concept of intentionality to the definition of bullying. Frequency of bullying – noted in the definition as repeatedly and over time – is included with “the intent...to exclude occasional non-serious negative actions that are directed against one student at one time and against another on a different occasion” (9). Repetition, then, becomes a key component of the definition of bullying. Finally, Olweus sets limits on the concept of bullying, noting that the term bullying is not “used when two students of approximately the same strength are fighting or quarrelling,” but rather only when there is an “imbalance in strength (an asymmetric power relationship)” (10). Power imbalance is the third component of the Olweus definition of bullying. Olweus separates bullying from other forms of peer

victimization using the power imbalance differential; he finds that victims of bullying “perceive significantly more threat and less control over their situation in addition to being more depressed, engaging in more wishful thinking, and seeking more social support” (Olweus 2010: 14) than other victims of peer aggression.

Since Olweus’ work in the 1970s and onward, researchers have expanded upon the attributes of bullying. In 2003, prolific bullying researchers Dorothy L. Espelage and Susan M. Swearer called for a consensus definition on bullying “the most challenging aspect of bullying prevention programming” (Espelage and Swearer 2003: 367). Twelve years later, Swearer echoes this lament (with Canadian bullying scholar, Shelley Hymel) that there “may be no single ‘gold standard’ for accuracy” in measuring bullying (2015: 294). There are operationalizations of bullying that focus on behaviors, attitudes, victims, perpetrators, and other indicators (Thomas et al 2015; Espelage and Swearer 2003), and those that focus on the various cultural settings of the research (Smith et al. 2013).

In addition to multiple definitions and ways of operationalizing bullying behavior, there are also differing means of collecting bullying victimization estimates (Thomas et al. 2015: 135; Espelage and Swearer 2003). These include observations (including unstructured and structured observations and interviews); teacher ratings (whereby teachers identify the bullies and victims); self-report (students self-nominate as bullies or victims); and instruments (booklets, inventories, scales, and other psychological tools).² By far, the most common means of collecting data for bullying victimization estimates is through questionnaires and surveys. Not only does this method protect student anonymity, it also allows for large-scale data collection and change over time. Using questionnaires, however, the burden of question interpretation falls to the respondent, which can lead to measurement error. As such, Cornell and Bandyopadhyay caution that "self-report measures are dependent on the student's understanding of the survey questions and his or her memory for events that may be unpleasant to recall" leading "some students...to inflate accounts of their experiences, while others may minimize or deny their involvement in bullying" (2010: 267).

Estimates of Bullying Victimization

The lack of uniformity in the definition of bullying contributes to the wide range of estimates on the prevalence of bullying. The three most commonly cited sources of national bullying estimates currently range widely in their reported rates of bullying. Although all three estimates are derived from self-reported surveys of youth, they use varying definitions of bullying and are therefore measuring different – though related – concepts³. The most extensive definitions – found on the Health Behavior in School-aged Children Survey (HBSC) and the Youth Risk Behavior Survey (YRBS) – outline for respondents the behaviors that constitute bullying (and those that do not), the power imbalance, and the repetitious nature of bullying. The School Crime Supplement to the National Crime Victimization Survey (SCS) provides the broadest definition, focused only on the victim’s feelings about others’ behaviors toward them. The differences in definition result in widely varying estimates. The SCS, with the broadest definition, has the highest national estimate of bullying, at 28 percent (Lessne and Harmkar 2013). The YRBS estimates that 20 percent of respondents have been bullied (National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention 2015), and the HBSC has the lowest estimate, at 11 percent (Iannotti 2013).

In 2014, the (CDC) reviewed the academic and applied literature on bullying and, in recognition of the plurality of definitions and measurements, published a uniform definition of the term to guide researchers in operationalizing the phenomenon. This new definition attempted to tie together the public health perspective (focused on unequal power and repetition), with the behavior and consequences focus of the educational and human services communities. The result was the following CDC uniform definition of bullying:

Bullying is any unwanted aggressive behavior(s) by another youth or group of youths who are not siblings or current dating partners that involves an observed or perceived power imbalance and is repeated multiple times or is highly likely to be repeated. Bullying may inflict harm or distress on the targeted youth including physical, psychological, social, or educational harm. (Gladden et al. 2014: 7)

² See Crothers and Levinson (2004).

³ Each of the three surveys also collects data on slightly different (but overlapping) populations.

Table 1: Sources of National Estimates of Bullying

Source	Population	Definition	Year	Estimate	Agency Sponsor
School Crime Supplement	12 to 18 year olds	Now I have some questions about what students do at school that make you feel bad or are hurtful to you. We often refer to this as being bullied. You may include events you told me about already. During this school year, has any student bullied you?	2011	28%	Department of Education, National Center for Education Statistics, and Department of Justice, Bureau of Justice Statistics
Youth Risk Behavior Survey	High school students	The next two questions ask about bullying. Bullying is when one or more students tease, threaten, spread rumors about, hit, shove, or hurt another student over and over again. It is not bullying when two students of about the same strength or power argue or fight or tease each other in a friendly way.	2011	20%	Centers for Disease Control, Division of Adolescent and School Health, National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention
Health Behaviors in School-age Children	11, 13, and 15 year olds	Here are some questions about bullying. We say a student is BEING BULLIED when another student, or a group of students, say or do nasty and unpleasant things to him or her. It is also bullying when a student is teased repeatedly in a way he or she does not like or when he or she is deliberately left out of things. But it is NOT BULLYING when two students of about the same strength or power argue or fight. It is also not bullying when a student is teased in a friendly and playful way.	2010	11%	World Health Organization, Child and Adolescent Health Research Unit

Aligning the SCS

A practical result of the publication of the CDC uniform definition is the need to update the surveys collecting bullying data to align with the revised definition. The School Crime Supplement (SCS) to the National Crime Victimization Survey (NCVS) is an annual survey of households in the United States. The SCS collects national-level data on students’ reports of school crime. The survey instrument was designed by the National Center for Education Statistics (NCES) and the Bureau of Justice Statistics (BJS). It asks students a number of questions about their experiences with and perceptions of crime and violence occurring inside their school, on school grounds, on the school bus, and from 2001 onward, going to or from school.

Additionally, the SCS includes questions about students’ schools and behavior, such as preventive measures used by the school, engagement in after-school activities, perceptions of school rules, weapons and gangs in school, hate-related words and graffiti in school, and others. The survey was conducted in 1989, 1995, 1999, and biennially since 1999; each year of data collection has been fielded by the Census Bureau. The SCS is a major source of national estimates of bullying in school and cyber-bullying anywhere.

The SCS questionnaire is administered after the NCVS to eligible persons in the sample. Eligibility includes those respondents ages 12 through 18 who are currently enrolled in a primary or secondary education program leading to a high school diploma or who were enrolled at some time during the school year of the interview, and did not exclusively receive their education through homeschooling during the school year⁴. All NCVS respondents aged 12 through 18 within NCVS households between January and June of the year of data collection are eligible to be screened for the SCS.

The SCS Measures Bullying

Over the years of data collection, the SCS has become a major source for national prevalence estimates of bullying in school and cyber-bullying anywhere. Respondents answer questions on bullying in each year of data collection. However, the wording of the questions has shifted from year to year, particularly between 2003 and 2005.

Beginning in 2005, the operationalization of bullying changed from a one-dimensional yes/no dichotomous question to a list of seven discrete bullying behaviors from which respondents are asked to choose. Selecting “yes” on any of these behaviors counts the respondent as being “bullied;” selecting “no” on all of the behaviors, or “no” on some behaviors and “don’t know” or “missing” on others counts the respondent as being “not bullied.” Note that respondents missing all data, or having “don’t know” for all seven indicators, are dropped from the bullying analyses (set as missing). Because of this dramatic shift in questions’ wording, the SCS bullying estimate trend line is truncated at the year 2005.

Even with the changes in the bullying question(s) over time, the 2013 SCS was not aligned with the CDC Uniform Definition of bullying published in 2014. While the SCS asked about aggressive behaviors that could inflict harm or distress, it did not specifically ask about a power imbalance, repetition (real or perceived), or harm caused by the behavior. As a result, the NCES determined that the SCS bullying questions would need to be redesigned to align with the CDC definition.

Technical Review Panel

On August 12-13, 2013, NCES convened a Technical Review Panel (TRP) Conference to discuss changes to the SCS, particularly the bullying questions. The panel consisted of 30 experts, ranging from independent contractors to academics, federal bureaucrats from the ED, DOJ, and the Census Bureau, and stakeholders from non-profit organizations and schools. The purpose of the TRP was to revisit the survey and examine the questions to determine which should change and which should be dropped (Zantal-Wiener and Lessne 2013: 2).

The experts at the TRP were frank about the lack of clarity in definitions of bullying from instrument to instrument, but also hesitant to settle on the intention or appropriate way of gathering bullying data. One expert said, hyperbolically, “We have had about 15 thousand definitions of bullying. There is no perfect definition, [but] the CDC definition is a great definition and a good starting point” (5). Another noted that by using the CDC definition of bullying, the intention of the question changed. Under the original wording, the intent of the question was the “degree” of bullying, but aligning with the new definition shifts the question’s intent to “instance” of incidences (15). By moving to the new definition, the “intent of the question [stem] is for students to self-identify as a bullied student” (8) more efficiently (that is, in fewer questions) rather than to describe the ways in which they may have been bullied.

⁴ Persons who have dropped out of school, have been expelled or suspended from school, or are temporarily absent from school for any other reason, such as illness or vacation, can complete the SCS as long as they have attended school at any time during the school year of the interview. Students who receive all of their education through homeschooling are not included past the screening questions, and those who receive part of their education through homeschooling are not included in my analyses.

Table 2: Bullying Question(s) for the School Crime Supplement Survey: 2001 to 2013

Year	Question(s)
2001	During the last 6 months have you been bullied at school? That is, has anyone picked on you a lot or tried to make you do things you didn't want to do like give them money?
2003	During the last 6 months, have you been bullied at school? That is, have any other students picked on you a lot or tried to make you do things you didn't want to do like give them money?
2005	During the last 6 months has any other student bullied you? That is, has another student... Made fun of, called names Spread rumors Threatened you Pushed, shoved, tripped Do things not wanted Excluded you Destroyed your property None of the above
2007	Now I have some questions about what students do at school that make you feel bad or are hurtful to you.
2009	We often refer to this as being bullied....During this school year, has any other student bullied you? That is,
2011	has another student...
2013	Made fun of, called names Spread rumors Threatened you Pushed, shoved, tripped Do things not wanted Excluded you Destroyed your property None of the above

The tension between identifying and describing bullying in the question was a conversation around which the experts had some debate. One suggested that the SCS “focuses on the behaviors and whether they were repeated” (20); another echoed this call saying that the question should be “behavioral” (5), while a third argued that “there must be a way to get at the severity or degrees of bullying” (6) rather than just counting instances. However, others noted that the SCS in particular provides national estimates, not necessarily a description of the bullying. Another expert said, “the intent of the [question] stem is for the student to self-identify as a bullied student” (8) rather than to describe the bullying behavior. Another agreed, saying that while “on the NCVS crime questions, we ask ‘what happened’,” this is “different than a self-report attribute-based system like the SCS” because “it would be harder for students to define particular incidents of bullying [and] to allow coding of each incident” since incidences are not usually discrete (4).

A major concern of the TRP was the loss of trend data in the national estimates of bullying, if the question wording was changed. Experts first debated whether or not the loss of trend data was worth aligning with the CDC definition; one said “that is the core question: is it ok to lose the data in favor of moving toward a uniform definition?” (9). Another pointed out that there is “a lot of room for improvement” on the bullying questions, but that researchers “would lose trend data and would be starting over” (9). One reminded the group that the SCS bullying question has a history of change – in 2005, the bullying question shifted from a yes/no question to a listing of bullying behaviors – and that the result was the rate of students identifying as bullied “jumped to 28 percent” because listing “the items [is] more concrete than trying to determine if you were bullied” with a yes/no question (12).

Along with the conversation about losing the bullying trend data, there was the concern that others might see the change as manipulative. One expert pointed out that since they could not provide trend data, they would “need to tell policymakers that we changed how we measured bullying so we don’t have trend data for another few years,” implying that this answer would not satisfy policymakers (9). Another pointed out that if the group of experts recommended changing the bullying question and the impact on the national estimates were dramatic, “the public might think that the Department of Education made the survey show that rates are going down” rather than understand that the change is a reflection of how the question is asked. Still these arguments were countered by one

expert who pointed out that “if we stay with the current question, we are asking from an exemption” from compliance with the new definition, and that the SCS “won’t be useful in new discussions” of bullying (9).

Although one expert called for an “immediate transition” to the new definition (12), even at this early stage of the redesign, some experts were advocating a split-ballot experiment, or at the very least, a pilot test. One called the new questions “definition 1.0” and called for the group to “put it out and test it” (5). Another suggested conducting a pilot survey where “half [of respondents] would include the old question stem and half the new stem” to see the changes in responses (10). In fact, such an experiment “need not be large to obtain an answer as to how response rates would be affected using the new language” (19).

Cognitive Testing

In response to the suggestions for rewording brought forth by the Technical Review Panel, four researchers from the Census Bureau’s Center for Survey Measurement (CSM) conducted 40 cognitive interviews. These respondents were split into two groups and were asked one of the two suggested new forms of the bullying question. For each group, respondents were both asked to “think out loud” while answering the questions and were also asked a series of follow up questions and probes based on their responses. While both versions were found to be effective, the results of the cognitive interviewing demonstrated that question re-wording would impact the national estimates on bullying.

For round one of the cognitive interviewing, respondents were asked a single yes/no question about bullying, and followed up with yes/no questions on bullying behavior. The behavior listed is taken from the SCS 2013 instrument. The text of this tested version is:

Round 1:

Q: Now I have some questions about what students do at school that makes you feel bad or is hurtful to you. We often refer to this as being bullied. You may include events you told me about already. During this school year, has any student bullied you?

A: Yes/No

Q: That is, has another student...

Made fun of you, called you names, or insulted you, in a hurtful way?

Yes/No

Spread rumors about you or tried to make others dislike you?

Yes/No

Threatened you with harm?

Yes/No

Pushed you, shoved you, tripped you, or spit on you?

Yes/No

Tried to make you do things you did not want to do, for example, give them money or other things?

Yes/No

Excluded you from activities on purpose?

Yes/No

Destroyed your property on purpose?

Yes/No

The objective of testing this version of the question was “to assess whether there were any disconnects between the answers to a general yes/no question on bullying and the specific incidents” described in the survey (Pascale et al. 2014: 14). Results of the cognitive interviewing for this version of the question demonstrated that the inclusion of specific behaviors may be increasing the counts of bullied students, such that “we had no instances of students saying ‘yes’ to the general question and then ‘no’ to the specific incidences” but “we found that some students said ‘no’ to the yes/no question, but then ‘yes’ to one or more incidents described” in the follow up question (14). Since SCS 2013 did not include the general yes/no question, but rather constructed a bullying variable based on responding yes to one or more of the listed behaviors, those students saying ‘no’ to the general question and ‘yes’ to a behavioral follow-up would have been counted as a bullied student under the 2013 SCS.

CSM researchers then tested an alternative version of the bullying questions. In this case, students were presented a single bullying question with the complete definition of bullying embedded in the question stem, including repetition and power imbalance. Specific behaviors were only presented as follow-up to those students who already identified as being bullied based on the yes/no single bullying question. The form of this question series was:

Round Two:

Q: Now I have some questions about bullying at school. Bullying happens when one or more students tease, threaten, spread rumors about, hit, shove, or hurt another student. It is not bullying when students of about the same strength or power argue or fight or tease each other in a friendly way. Bullies are usually stronger, or have more friends or money, or some other power over the student being bullied. Usually, bullying happens over and over, or the student being bullied thinks it might happen over and over. By this definition, have you been bullied at school, by another student, this school year?

A: Yes/No

IF YES

Q: Was any of the bullying verbal – that is, did it involve making fun of you, calling you names, or spreading rumors about you?

A: Yes/No

Q: Was any of the bullying physical – that is, did it involve hitting, shoving, tripping, or physically hurting you in some way?

A: Yes/No

Q: Was any of the bullying social – that is, did it involve ignoring you or excluding you from activities on purpose in order to hurt you?

A: Yes/No

Results of the cognitive interviewing on this version of the questions demonstrated that the question series seemed to work as intended. Those who answered yes to the general question then gave examples that seemed to fit the prescribed definition; those who answered no to the general question indicated that “nothing in their experience seemed to meet the definition of bullying” (Pascale, et al. 2014: 16). Researchers did point out, however, that while the question did not seem problematic, it is “quite long” and that “respondents are likely artificially attentive when being asked questions in a face-to-face lab setting” compared to out in the field via CATI or CAPI interviews; however, they also noted that “it does seem that the clear definition of bullying helps students decide how to answer the question” (16).

Once the recommendations from the Technical Review Panel and the results of the cognitive interviewing were analyzed, NCES moved forward with aligning the 2015 SCS bullying question series to the uniform definition provided by the CDC. To align the bullying definition for the SCS while still retaining the ability to compare estimates across years, NCES would need to collect the new data in a way that allows for a bridge year to the old definition. The best way to collect these data was through the use of a split-ballot experiment imbedded in the instrument. The split-ballot experiment is a way of randomly assigning sampled respondents into two or more groups and either administering the established survey instrument (the “control” group), or the new survey instrument (the “experimental” group).

However, the split-ballot experiment has methodological limitations, too. It can, for example, “manipulate only a single factor, and the manipulated factor [can] assume only one of two values” (Sniderman and Grob 1996: 379). At the same time, it is only useful “to identify method-driven variance,” and is not necessarily theoretically driven as we cannot ask why changing a survey question solicits a different response, we can only say that it does (380-381).

Even noting the limitations, NCES determined that having the ability to continue the bullying trend line over time was the most important aim of the redesign next to alignment with the uniform definition, and the split-ballot experiment was the most efficient way to continue the trend. To accomplish this aim, then, half of the respondents to SCS 2015 were to be randomly assigned to either the control (established form) or experimental (new form) sample groups. The control and experimental groups shared the same questionnaire; only the approach to the bullying question varied. In addition, a third group – the “control” plus two follow-up questions asking about power imbalance and repetition – would act as an “in between” for further refinement of the bullying question. All respondents randomly selected to receive the control question form also received the two follow-up questions.

Table 3: Control and Experimental Question Series for SCS 2015

Control	Experimental	Rationale
<p>Now I have some questions about what students do at school that makes you feel bad or are hurtful to you. We often refer to this as being bullied. You may include events you told me about already. During this school year, has any student bullied you? That is, has another student....</p> <p>Made fun of you, called you names or insulted you, in a hurtful way?</p> <p>Spread rumors about you or tried to make others dislike you?</p> <p>Threatened you with harm?</p> <p>Pushed you, shoved you, tripped you, or spit on you?</p> <p>Excluded you from activities on purpose?</p> <p>Destroyed your property on purpose?</p> <p>Tried to make you do things you did not want to do, for example, give them money or other things?</p> <p>Excluded you from activities on purpose?</p>	<p>Now I have some questions about bullying at school. Bullying happens when one or more students tease, threaten, spread rumors about, hit, shove or hurt another student. It is not bullying when students of about the same strength or power argue or fight or tease each other in a friendly way. Bullies are usually stronger, or have more friends or money, or some other power over the student being bullied. Usually, bullying happens over and over, or the student being bullied thinks it might happen over and over.</p> <p>By this definition, have you been bullied at school, by another student, this school year?</p>	<p>This is the question that will determine the student's bullying status.</p>
	<p>Was any of the bullying verbal - that is, did it involve making fun of you, calling you names, or spreading rumors about you?</p> <p>Was any of the bullying physical - that is, did it involve hitting, shoving, tripping, or physically hurting you in some way, or the threat of hurting you in some way?</p> <p>Was any of the bullying social - that is, did it involve ignoring you or excluding you from activities on purpose in order to hurt you?</p>	<p>Collects data on the type of behaviors involved in the bullying incidents.</p>
<p>When you were bullied this year, did it happen over and over, or were you afraid it would happen over and over?</p>		<p>Answers the repetition definition requirement.</p>
<p>When you were bullied this school year, were you ever bullied by someone who had more power or strength than you? This could be because the person was bigger than you, was more popular, had more money, or had more power than you in another way?</p>		<p>Answers the power imbalance definition requirement</p>
<p>You just indicated that someone had bullied you during this school year. Thinking about all of the ways in which you were bullied, how often did all of those things happen?</p>	<p>You just indicated that someone had bullied you during this school year. Thinking about all of the ways in which you were bullied, how often did all of those things happen?</p>	<p>Experimental and control group realign for the remainder of the survey.</p>

Ultimately, the review of the SCS instrument resulted in more than just changes to the bullying questions. A number of additional survey items were revised, added, or deleted, including⁵:

- Redesigned key bullying questions
- Reduced net number of survey items by 12
- Revised wording on 16 questions for clarity and updates to current terminology
- Added/revised instructions for respondents
- Renumbered all items to aid field representatives and researchers in tracking related sequences of items

⁵ See Lessne and Cidade 2017 for a more detailed accounting of the additional changes to the SCS 2015 instrument.

Note, too, that a third, mid-point bullying measure is embedded in this construction. Any respondent in the “control” group that responded yes to being bullied also received two follow-up questions that measure repetition and power imbalance.

Repetition:

When you were bullied this year, did it happen over and over, or were you afraid it would happen over and over?

Power Imbalance:

When you were bullied this school year, were you ever bullied by someone who had more power or strength than you? This could be because the person was bigger than you, was more popular, had more money, or had more power than you in another way?

In this way, the SCS could capture bullying using the historic question, but pare down the responses to those who meet the repetition and power imbalance aspects of the CDC definition. This construction was labeled “control plus follow-up” and is included in the results below.

In consultation with the Demographic Statistical Methods Division of the Census Bureau, NCES determined that a true split ballot (50/50 random assignment) would produce estimates in differences in the bullying rate of 10 percent as significant.

Analyzing the Split-Half

The first step to analyzing the results of the split-half experiment is to account for any differences in survey response between the two samples. Before this work could begin, ineligible cases were dropped from the data file. This included dropping those who did not attend school at all, those who were homeschooled (for the full and partial school year), and those who were not in grades six through 12. These parameters are the ones used by the NCES in previous analyses of the SCS.

In January to June of 2015, there were 57,227 households eligible to complete the NCVS. The SCS questionnaire is administered after the NCVS to eligible respondents in the sample. Among those households participating in the NCVS, there were 9,372 respondents ages 12-18 who were eligible to complete the SCS in 2015. Among the 9,372 household members age 12-18, version 1 of the survey form (control) was assigned to 4,663 respondents (49.7%) and version 2 was assigned to 4,709 respondents (50.3%). Of the 9,372 age-eligible individuals in NCVS households, 5,469 completed the NCVS survey and were interviewed for the SCS. Once the responses were filtered by eligibility criteria described earlier, a total of 4,767 completed the survey, of whom, 2,317 completed version 1 (control) and 2,386 completed version 2 (experimental).

This “cleaned” data file, then, could produce the estimates of the three measures of bullying, control, control plus follow-up questions, and experimental. Unsurprisingly, including the definition of bullying on the survey instrument lead to lower prevalence estimates in bullying victimization.⁶ Asking about bullying victimization by behavior as in previous years, 20.8% of students in grades 6 through 12 report being bullied. Further refining the analysis to those students who reported being bullied and said yes to both power imbalance and repetition follow-up questions, the percentage of students satisfying the CDC definition of bullying is 4.5%. Similarly, for the experimental group, when first given the CDC definition of bullying and then asked a dichotomous question about bullying victimization, just 8.1% of students in grades 6 through 12 report being bullied.⁷

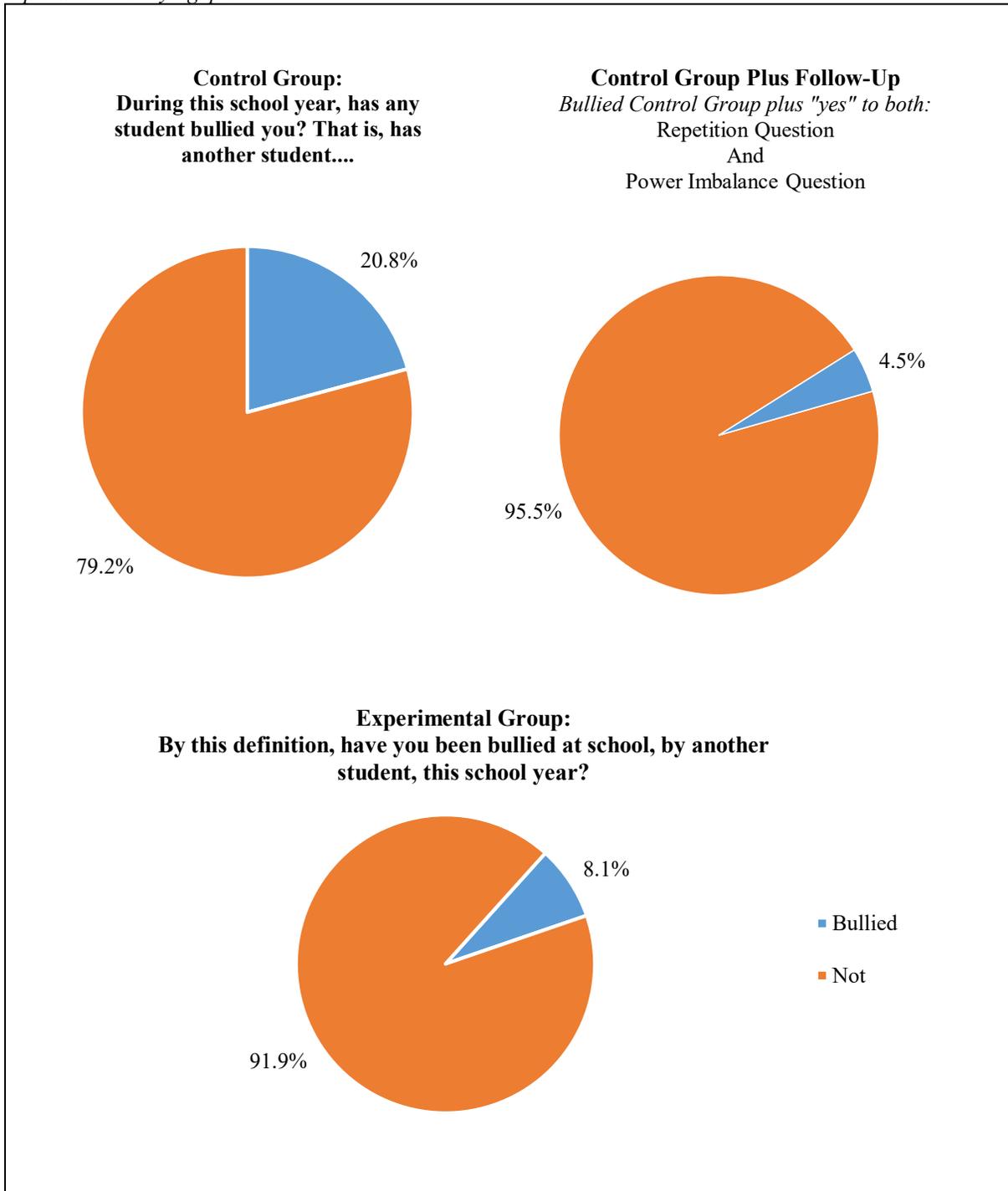
⁶ See Solberg and Olweus 2003 for more on including definitions

⁷ Note: Both the experimental and control estimates are weighted estimates. To apply the weight to each randomized half of the sample, take the overall SCS person weight and divide by 2 before applying to the data. All data were analyzed using SPSS Complex Samples to handle both the effects of the sample design and the weighting effects. For more on the weighting scheme of the 2015 administration of the SCS, including detail on weighting and the split-ballot, see Lessne and Cidade 2017: 12.

Table 4. Ineligible Cases Dropped from Analysis

	Control		Experimental		Overall	
	Unweighted	Weighted	Unweighted	Weighted	Unweighted	Weighted
Total interviews	4,663	----	4,709	----	9,372	----
Incomplete interviews	1,969	----	1,934	----	3,903	----
Total valid interviews	2,694	14,231,000	2,775	14,786,000	5,469	29,017,000
Did not attend school	161	888,000	145	803,000	306	1,691,000
Total valid interviews and attended school	2,533	13,344,000	2,630	13,982,000	5,163	27,326,000
Homeschooled at any point in the school year	65	326,000	92	501,000	157	833,000
Total valid interviews, attended school, no homeschooling	2,468	13,018,000	2,539	13,488,000	5,007	26,506,000
Not in grades 6 through 12	151	707,000	153	835,000	304	1,542,000
Total valid interviews, attended school, no homeschooling, in grades 6 to 12	2,317	12,311,000	2,386	12,653,000	4,703	24,964,000

Figure 1: Percentage of students in grades 6 through 12 who reported being bullied or not bullied, by control or experimental bullying question



Lessons Learned

During initial discussions of item development and planning of the split-half experiment, a common opinion among the three agencies involved, as well as support contractors, was one in support of the new, experimental version of the item. While there was concern about the length of the item, this version was popular as it provided all components of the CDC definition up front and was assumed to yield a more accurate estimate of bullying victimization. However, once the results of the 2015 SCS randomized split-ballot experiment were analyzed, all

three federal agencies – NCES, BJS, and Census – determined that the estimates were unstable and that continued work on the items was necessary.

Several factors influenced this decision. First, looking at the control version, since the trend estimate continued to slightly drop (2013: 21.5% vs 2015: 20.8%), researchers could be fairly confident in the reliability of the item wording. The wording for the trend estimate control group had not changed from prior administrations, and the estimates stayed comparable. However, the drastic drop for the control when taking into account the two components of the CDC definition (repetition and power imbalance; 20.8% vs 4.8%) that had not been previously collected on the SCS indicated potential measurement error and comprehension issues. At the same time, the experimental question dropped bullying victimization at a rate much higher than expected (8.1% reported bullying victimization). Again, we could not be confident in the cause of this drop and therefore knew continued work would be necessary if the item was retained for future administrations. Knowing both options would require continued revisions, the mode of data collection (CAPI) had to be taken into consideration. For the SCS, interviewers read the items to the respondents. The length of the experimental version required an inordinate amount of attention and retention in order to accurately answer the question.

Due to these concerns, the 2017 administration of the SCS retained the control version of the bullying victimization question, including the follow-up questions on repetition and power imbalance. Continuing with the control question will maintain a decade-long trend of bullying data, while including two follow-up questions satisfies the CDC Uniform Definition. The SCS data are critical to academic and applied bullying and harassment researchers, and collecting these data in a way that is valid but also aligned with changes in the field is paramount. Further, this approach sets the foundation for future work involving collecting detailed information on student experiences in schools and allowing researchers to investigate the various pieces of the uniform definition separately, or as a whole. This is a unique approach to collecting these data at the federal level.

The collaborative work between NCES, BJS, and Census has been crucial in determining the best way to approach incorporating the components of the uniform definition, as well as determining the best path moving forward in future administrations of the SCS. Given the supplemental nature of the SCS collection, BJS is able to provide first-hand knowledge of their experiences with other supplements to the NCVS. When BJS developed the Stalking Supplement to the NCVS, researchers noted that some respondents reacted negatively to the word “stalking” on the survey. This work is informing current cognitive testing for the 2019 SCS bullying questions. For example, for the 2019 SCS, researchers are cognitively testing the entire bullying section on the SCS with the term “bullied” removed from all items. Removing this term focuses respondents on their victimization experiences, followed by detailed questions that measure repetition and power imbalance, without defining the term at the onset. In this way, the approach is similar to what was utilized in the development of the Stalking Supplement to the NCVS. The terms “bullying” and “stalking” are similar in that they typically have negative connotations, people have their own predetermined definitions, and some people may be reluctant to admit experiencing bullying or stalking. Initial findings of the cognitive testing of the 2019 items have shown respondents favoring this approach to collecting this information, with some respondents admitting they likely would have responded in the negative had they been specifically asked if they had experienced being bullied.

The collaboration among these agencies continues as the Census Bureau is the data collector for the NCVS, and their Center for Survey Measurement (CSM) was involved in the item development and cognitive testing of both the SCS and the Stalking Supplement. This overlap has allowed for quick comparisons between the supplements and for SCS to essentially “piggy-back” off of the lessons learned for the Stalking Supplement testing. That seamless collaboration, along with overlapping content expertise from ED and BJS, has led to the continued refinement of these important measures.

While having three agencies involved in determining major changes in item wording could result in a “too many cooks in the kitchen” scenario, this has not been the case for aligning the SCS bullying questions to the CDC Uniform Definition. The expertise provided from all three agencies was instrumental in developing an action plan to collecting data to reflect the CDC definition. This relationship furthers work towards collecting a wealth of data that will continue to inform researchers and policy-makers of students’ experiences in school.

References:

- Cornell, Dewey G. and Sharmila Bandyopadhyay. 2010. "The Assessment of Bullying." Pp. 265–76 in *Handbook of Bullying in Schools: An International Perspective*, edited by S. R. Jimerson, S. M. Swearer, and D. L. Espelage. New York: Routledge.
- Crothers, Laura M. and Edward M. Levinson. 2004. "Assessment of Bullying: A Review of Methods and Instruments." *Journal of Counseling and Development* 82:496–503.
- Espelage, Dorothy L. and Susan M. Swearer. 2003. "Research on School Bullying and Victimization: What Have We Learned and Where Do We Go from Here?" *School Psychology Review* 32(3):365–83.
- Gladden, R. Matthew, Alana Vivolo-Kantor, Merle E. Hamburger, and Corey D. Lumpkin. 2014. *Bullying Surveillance Among Youths: Uniform Definitions for Public Health and Recommended Data Elements, Version 1.0*. Atlanta, GA: National Center for Injury Prevention and Control, Centers for Disease Control and Prevention and U.S. Department of Education.
- Hymel, Shelley and Susan M. Swearer. 2015. "Four Decades of Research on School Bullying." *American Psychologist* 70(4):293–99.
- Iannotti, R. J. 2012. *Health Behavior in School-Aged Children (HBSC), 2005-2006*. Ann Arbor, MI: Inter-university Consortium for Political and Social Research.
- Lee, Jesse. 2011. 'President Obama and the First Lady at the White House Conference on Bullying Prevention'. *The White House*. Retrieved January 14, 2018. (<https://obamawhitehouse.archives.gov/blog/2011/03/10/president-obama-first-lady-white-house-conference-bullying-prevention>).
- Lessne, Deborah S. and Sayali Harmalkar. 2013. *Student Reports of Bullying and Cyber-Bullying: Results from the 2011 School Crime Supplement to the National Crime Victimization Survey*. Washington, DC: U.S. Department of Education National Center for Education Statistics.
- Lessne, Deborah S. and Melissa A. Cidade. 2017. *Split-Half Administration of the 2015 School Crime Supplement to the National Crime Victimization Survey*. Washington, DC: U.S. Department of Education National Center for Education Statistics. Retrieved (<https://nces.ed.gov/pubs2017/2017004.pdf>).
- McCallion, G. & Feder, J. (2013). *Student Bullying: Overview of Research, Federal Initiatives, and Legal Issues*. Washington, DC: Congressional Research Service.
- National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention. 2015. *Trends in the Prevalence of Behaviors That Contribute to Violence on School Property National YRBS: 1991-2015*. Atlanta, GA: Centers for Disease Control and Prevention.
- Olweus, Daniel. 1978. *Aggression in the Schools: Bullies and Whipping Boys*. New York: Hemisphere Pub. Corp.
- Olweus, Dan. 1993. *Bullying at School: What We Know and What We Can Do*. Cambridge, MA: Blackwell Publishers.
- Olweus, Dan. 2013. "School Bullying: Development and Some Important Challenges." *Annual Review of Clinical Psychology* 9:751–80.
- Olweus, Dan. 2010. "Understanding and Researching Bullying: Some Critical Issues." Pp. 9–36 in *Handbook of Bullying in Schools: An International Perspective*, edited by S. R. Jimerson, S. M. Swearer, and D. L. Espelage. New York: Routledge.

Pascale, Joanne, Mikelyn Meyers, Mandi Martinez, and Marissa Fond. 2014. *National Crime Victimization Survey School Crime Supplement Cognitive Testing of Questions on Bullying*. Washington, DC: US Census Bureau, Center for Survey Measurement.

Safe Schools Improvement Act of 2011, S.506, 112th Congress (2011).

Smith, Peter K., Christina del Barrio, and Robert S. Tokunaga. 2013. "Definitions of Bullying and Cyberbullying: How Useful Are the Terms?" Pp. 26–40 in *Principles of Cyberbullying Research: Definitions, Measures, and Methodology*, edited by S. Bauman, D. Cross, and J. Walker. New York: Routledge.

Sniderman, Paul and Douglas Grob. 1996. "Innovations in Experimental Design in Attitude Surveys." *Annual Review of Sociology* 22:377–99.

Solberg, Mona and Dan Olweus. 2003. "Prevalence Estimation of School Bullying with the Olweus Bully/Victim Questionnaire." *Aggressive Behavior* 29:239–68.

Thomas, Hannah J., Jason P. Connor, and James G. Scott. 2015. "Integrating Traditional Bullying and Cyberbullying: Challenges of Definition and Measurement in Adolescents - a Review." *Education Psychology Review* 27:135–52.

U.S. Department of Education. 2014. *U.S. Department of Education Strategic Plan for Fiscal Years 2014-2018*. U.S. Department of Education: Washington, DC.

U.S. Department of Health and Human Services. 2014. *U.S. Department of Health and Human Services: Strategic Plan 2014-2018*. U.S. Department of Health and Human Services: Washington, DC.

U.S. Department of Justice. 2014. *U.S. Department of Justice Fiscal Years 2014-2018 Strategic Plan*. U.S. Department of Justice: Washington, DC.

Zantal-Wiener, Kathy and Deborah Lessne. 2013. *ESSN Crime and Safety Support: School Crime Supplement (SCS) to the National Crime Victimization Survey (NCVS) 2013 - Minutes of the Technical Review Panel*. Washington, DC: National Center for Education Statistics.