Memorandum

United States Department of Education
Institute of Education Sciences
National Center for Education Statistics

DATE: December 18, 2015

TO: Robert Sivinski, OMB

THROUGH: Kashka Kubzdela, OMB Liaison, NCES

FROM: Carolyn Fidelman, MGLS:2017 Project Officer, NCES

SUBJECT: Middle Grades Longitudinal Study of 2017-18 (MGLS:2017) Recruitment for 2017
Operational Field Test (1850-0911 v.6) Responses to OMB Passback

This memorandum provides response to the OMB request for additional information regarding sampling plans for the recruitment package for the Middle Grades Longitudinal Study of 2017-18 (MGLS:2017) Operational Field Test.

OMB Comment 1: Please provide more detail on the sample design, including a description of the frame, a definition of the measure of size, and the PPS sample selection process at each stage.

Associated text (B.1): The MGLS:2017 OFT will employ a multi-stage probability proportional to size sampling design, with schools selected in one stage, and then students selected within schools.

OMB Comment 2: Does the certainty come from the PPS design or from external factors? If it's not PPS, how may certainty schools are there and what percent of the sample do they represent? What's the resulting loss in sample efficiency?

Associated text (B.1): Schools will be selected using probability proportional to size sampling within school sampling strata, though some schools may be selected with certainty.

OMB Comment 3: What are these rates and how were they decided?

Associated text (B.1): The sampling rate, r_j, equals the number of students to sample from the jth category divided by the number of students in the jth category across all schools in the sampling frame.

OMB Comment 4: What's the sample allocation amongst the four strata? Is it equal? Is this design identical to the full-scale collection design?

Associated text (B.1): Within participating schools, students will be stratified into the four student categories defined above and a simple random sample of students will be selected from each sampling stratum.

NCES: A portion of section B.1 has been revised as follows (with the remainder of section B.1 staying unchanged):

B.1 Respondent Universe

The MGLS:2017 Operational Field Test (OFT) will be conducted during the 2016-17 school year, with data collection scheduled to begin in January 2017. The OFT will be conducted in ten geographic locations that adequately simulate the diversity of the 50 States and the District of Columbia. The OFT may include up to two quasi-nationally representative samples of students in the United States. The Sample 1 universe will include students enrolled in grade 6 and attending general education schools, while Sample 2, if drawn, would include students in general education schools in three focal disability categories (autism, specific

learning disability, and emotional disturbance) who are enrolled in grade 6 or are of equivalent age in an ungraded classroom. Sample 2 will be used if Sample 1 proves insufficient to meet the target student sample yields for the three focal disability categories.

The MGLS:2017 OFT will employ a multi-stage sampling design, with schools selected in one stage, and then students selected within schools. Schools will be selected using probability proportional to size sampling within school sampling strata, with some schools selected with certainty in order to ensure that all desired types of schools are included (see below).

Students will be selected using simple random sampling within student sampling strata within schools. The school frame will be constructed from the 2013-2014 Common Core of Data (CCD 2013-14) and the 2013-2014 Private School Universe Survey (PSS 2013-14) and will include 3,301 schools that report offering sixth-grade instruction and are located within ten metropolitan statistical areas (MSAs). The following types of schools will be excluded from the sampling frame:

- Department of Defense Education Activity schools and Bureau of Indian Education schools,
- schools for juvenile delinquents,
- schools that report no[1] sixth-grade enrollment and do not report to *EDFacts*,
- schools that report no sixth-grade enrollment and report no[2] students between the ages of 11 and 13 in the three focal disability groups,
- special education schools, and
- schools included in the Item Validation Field Test (IVFT) field test.

Schools will be stratified by MSA and by a High/Low Prevalence designation derived from the total number of students in the three focal disability groups. Schools will be classified as High Prevalence if the total number of students in the three focal disability groups exceeds the 95th percentile for that total across all[3] schools not just those schools in the ten MSAs used for the OFT sampling. One hundred and twenty five schools will be sampled and 103 of the 125 schools will be selected for initial data collection. The 22 schools not selected for initial data collection will be used as a reserve sample in the event that a higher than expected proportion of schools decline to participate or are found to be ineligible for the OFT.

The allocation of the 125 schools to the school sampling strata proceeded by first determining an allocation of 103 schools across the school sampling strata in order to have approximately half of the sample consist of high prevalence schools, half of the sample consist of low prevalence schools, and have approximately ten sample schools in each MSA. Twenty two additional reserve schools were distributed across the school sampling strata in order to try and preserve these goals as well. The school strata and a sample allocation that meets these goals are shown in table 1.

---

[1] Sixth-grade enrollment is reported as 0 or was not reported to the CCD 2013-14 or PS 2013-14.
[2] A school reports zero students or does not report the number of students in any of the three focal disability groups.
[3] The reason for this is to mimic the cutoff that is planned to be used for the national data collection in order to gauge the degree to which high prevalence schools agree to participate.

Table 1. OFT School Sample Allocation

School Region	Prevalence	School Frame Count	School Sample Allocation (103)	30 Certainty Private, Town, and Rural Schools	Certainty Schools Due to 100 Percent Sampling	Non Certainty Schools Among 103 Schools	Reserve Sample	PPS Sample of 78	School Sample Size (125)
A	High Prevalence	76	7	2	0	5	2	7	9
A	Low Prevalence	395	5	3	0	2	1	3	6
B	High Prevalence	22	7	0	0	7	2	9	9
B	Low Prevalence	293	5	3	0	2	1	3	6
C	High Prevalence	1	1	0	1	0	0	0	1
C	Low Prevalence	87	6	1	0	5	2	7	8
D	High Prevalence	5	5	0	5	0	0	0	5
D	Low Prevalence	170	5	3	0	2	1	3	6
E	High Prevalence	5	5	0	5	0	0	0	5
E	Low Prevalence	305	5	3	0	2	1	3	6
F	High Prevalence	40	7	0	0	7	2	9	9
F	Low Prevalence	566	5	3	0	2	1	3	6
G	High Prevalence	14	7	0	0	7	2	9	9
G	Low Prevalence	572	5	3	0	2	1	3	6
H	High Prevalence	12	7	0	0	7	2	9	9
H	Low Prevalence	497	5	3	0	2	1	3	6
I	High Prevalence	4	4	0	4	0	0	0	4
I	Low Prevalence	148	5	3	0	2	1	3	6
J	High Prevalence	2	2	0	2	0	0	0	2
J	Low Prevalence	87	5	3	0	2	2	4	7
Total		3,301	103	30	17	56	22	78	125

The first step in the sampling process involves the selection of thirty schools with certainty in order to ensure that at least ten private schools, ten schools in rural areas, and ten schools in towns are included in the sample of 103 schools to be initially fielded for data collection. The allocation of these thirty schools to the twenty school sampling strata is provided in table 1. The second step in the sampling process requires identifying those school sampling strata where the sample allocation of the 125 schools equals the number of schools in the frame. As shown in table 1, there are five sampling strata where the sample size equals the number of schools in the frame and there are seventeen schools in these five strata. All seventeen schools will be selected with certainty. Forty seven of the 103 schools, or approximately 46 percent of the sample, will be selected with certainty. The third step in the sampling process involves selecting 78 schools using stratified probability proportional to size sampling. The distribution of these 78 schools across the school sampling strata is provided in table 1. The fourth step in the sampling process entails selecting a simple random sample, within school strata, of the 78 schools selected using probability proportional to size in order to select 56 of the 78 schools for inclusion in the initial set of 103 schools released for data collection. The sample allocation of these 56 schools is shown in table 1.

The size measure used for the probability proportional to size selection of 78 schools will be constructed using the overall sampling rates for students in the following four student categories:

- Autism,
- Emotional Disturbance,
- Specific Learning Disability, and
- Other

combined with the total number of students in each of those categories at a given school. In other words, the size measure for a given school (i) may be written as follows:

$$S_i = \sum_{j=1}^{4} r_j N_{i,j}$$

Where r_j is the sampling rate for the jth student category and $N_{i,j}$ is the number of students in the jth category within school *i*. The sampling rate, r_j, equals the number of students to sample from the jth category divided by the number of students in the jth category across all schools in the sampling frame. The sampling rate for autistic students equals 188/7,706 (.024), the sampling rate for students with emotional disturbance equals 188/5,115 (.037), the sampling rate for students with specific learning disability equals 188/42,603 (.004), and the sampling rate for other students equals 1,188/280,424 (.004.) The denominator of a given rate for a given domain corresponds to the number of students in that domain across all schools in the sampling frame. The numerator of a given rate equals the number of students required to be sampled in order to achieve 1,120 participating students including 120 autistic students, 120 students with emotional disturbance, 120 students with specific learning disability, and 760 other students.

The unequal weighting effect is designed to be one within each of the four student domains (autism, specific learning disability, emotional disturbance, and other) within each school stratum among the schools selected using probability proportional to size sampling. A design effect of one means that the precision of an estimate is equivalent to the precision of an estimate derived via simple random sampling. The degree to which certainty schools will reduce the precision of estimates depends upon the degree to which certainty schools participate, the degree to which students in all schools participate, the degree to which student enrollment counts match expected counts, and the set of non-certainty schools that are sampled.

Within participating schools, students will be stratified into the four student categories defined above and a simple random sample of students will be selected from each student sampling stratum. The number of students sampled per student stratum will vary by school because the within school student allocation to strata depends upon the number of students in each of the four student sampling strata. The process of determining the student sample allocation follows the procedure outlined in section 2 of Folsom et al (1987[4].) The number of sampled students per student domain will generally not be equal and will vary across schools. The process outlined in Folsom et al. will also be followed for the full-scale collection. Approximately 34 students will be sampled from each of the anticipated 50 participating schools.

For OFT Sample 1, schools will be selected for recruitment from the CCD 2013-14 to develop the public school list and from the PSS 2013-14 to develop the private school list. In Sample 1, once schools are selected and recruited, students enrolled in grade 6 will be selected from student rosters that schools will be asked to provide. For OFT Sample 2, districts would be contacted to provide rosters of students enrolled in grade 6 or of equivalent age in an ungraded classroom along with their disability codes. Students in the autism, specific learning disability, and emotional disturbance categories would be sampled from the provided rosters, and their schools would be contacted for recruitment of the specified students. Alternately, if districts prefer, they may identify a school or schools in their district which contain grade 6 students in one or more of the focal disability groups and MGLS:2017 staff will invite the school(s) to participate and request a student list with disability codes from the school(s).

---

[4] Folsom, R.E., Potter, F.J., and Williams, S.R. (1987). Notes on a Composite Size Measure for Self-Weighting Samples in Multiple Domains. Proceedings of the Section on Survey Research Methods of the American Statistical Association, 792-796.