Memo

Date: January 28, 2010

To: Gail Mulligan

From: Alberto Sorongon, Christine Nord, Karen Tourangeau

Subject: Report of Fall 2009 Field Test results for Science Academic Rating Scale (ARS)

In addition to the direct child assessments, the Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K) included Academic Rating Scales (ARS) in language and literacy, mathematics, and science (beginning in grade 3) completed by the children's teachers. The ARS were designed both to overlap and to augment the information gathered through the direct cognitive assessments. Most importantly, the ARS included items designed to measure both the process and products of children's learning in school, whereas the direct cognitive battery was more limited. Because of time and space limitations, the direct cognitive battery was less able to measure the process of children's thinking, including the strategies they use to read, solve mathematical problems, or investigate a scientific phenomenon. The ARS were successfully collected in the ECLS-K and will be repeated in the ECLS-K:2011.

In fall 2009, a field test was conducted to test the feasibility of including a Science Academic Rating Scale (ARS) that could be collected before grade 3. As with the ARS collected in the ECLS-K, the new Science ARS would complement the information collected through the direct child assessment of science knowledge. This memo describes the collection of the Science ARS during the field test, the performance of the field-tested Science ARS, and our recommendations for the national data collection.

 The field tested Science ARS questions focused on topic areas in science that are typically a part of the curriculum in the kindergarten, first, and second grade levels, with examples of relevant skills for each topic area. In order to capture the range of abilities children may have at each grade level, the examples of skills for each topic area included some high-level skills. For each item, teachers were asked to rate the child on the following **five-point scale** which reflects the degree to which a child has acquired and demonstrates the targeted skills, knowledge, and behaviors (see exhibit 1).

Exhibit 1.   Response scale for Science Academic Rating Scale (ARS)

Response Anchor	Definition	Value
Not yet	Child has not yet demonstrated skill, knowledge, or behavior.	1
Beginning	Child is just beginning to demonstrate skill, knowledge, or behavior but does so very inconsistently.	2
In progress	Child demonstrates skill, knowledge, or behavior with some regularity but varies in level of competence.	3
Intermediate	Child demonstrates skill, knowledge, or behavior with increasing regularity and average competence but is not completely proficient.	4
Proficient	Child demonstrates skill, knowledge, or behavior competently and consistently.	5
Not Applicable or Skill Not Yet Taught	Skill, knowledge, or behavior has not been introduced in classroom setting.	--

If the skill, knowledge, or behavior had been introduced in the classroom, teachers were instructed to rate only the child's **current** achievement. If the skill, knowledge, or behavior had not been introduced in the classroom setting, teachers were instructed to select "*Not Applicable or Skill Not Yet Taught.*"

In each school participating in the field test, teachers of children in kindergarten, first grade, and second grade were each sent a packet of 5 Science ARS forms. They were asked to choose 5 children from their class and rate their science skills and knowledge using the forms. Specifically, they were asked to rate their highest achieving student, their lowest achieving student, and three with average achievement. They were instructed not to indicate on the form who these children were, merely to select children who match the criteria and think about their skills and knowledge in science when completing the forms. Teachers indicated the child's ability level of the child on each form. As a result, Science ARS field test data were collected on children with different abilities.

Table 1 shows the number of teachers that participated in the field test. In total, we had 1,239 completed Science ARS form from 249 teachers.

Table 1.   Number of teachers participating in fall2009 field test of Science Academic Rating Scale (ARS) and the number of ratings forms completed by grade

	# of Participating Teachers	# of completed forms
Kindergarten	86	426
First Grade	85	423
Second Grade	78	390

Total	249	1,239

We reviewed the performance of the Science ARS in the field test separately by grade. First, we examined the distribution of scores overall and by ability level. We also examined its internal-consistency reliability by grade, and examine means scores by ability level. We then compared the performance of the Science ARS with that of the ARS instruments previously fielded in ECLS-K.

Kindergarten

Of the 426 Science ARS forms completed by kindergarten teachers, valid achievement level data were obtained for 423 forms. Eighty-five forms (20.1%) were for children who were the teachers' highest achieving student; 88 forms (20.8%) were for children who were the teachers' lowest achieving student; and 250 (59.1%) were for children with average achievement. These results indicate that teachers did not experience difficulties completing the form and that they followed the overall instructions regarding student ability levels.

Means and variances of kindergarten Science ARS item data show good variation, with no floor or ceiling effects. For each item, the full range of possible item values (i.e., 1-5) were used by kindergarten teachers. Table 2 shows the average item scores across the completed kindergarten forms, as well as the percent with the highest possible rating (5), the percent with the lowest possible rating (1), and the percent where "Not Applicable Or Skill Not Yet Taught" was selected for the item.

Table 2. Average item scores across the completed kindergarten Science ARS forms and percent where the highest possible rating, the lowest possible rating, and "Not Applicable Or Skill Not Yet Taught" was selected.

	Average Item Score	Percent Highest Possible Rating (5)	Percent Lowest Possible Rating (1)	Percent "Not Applicable Or Skill Not Yet Taught"
K_1. **Uses his/her senses to explore and observe –** for example, observes and notes the habits of classroom pets; identifies environmental sounds; or describes the differences in clay before and after water is added	3.5	23.7	2.6	5.9%
K_2. **Forms explanations based on observations and explorations –** for example, describes or draws the conditions (water, soil, sun) that help a plant grow; or explains that a block will slide more quickly down a steeper slope.	3.2	15.8	8.7	10.4%
K_3. **Classifies and compares living and non-living things in different ways** – for example, classifies objects according to "things that are alive and not alive," or "things that fly and things that crawl," or "plants and animals."	3.6	23.4	8.7	13.2%
K_4. **Makes logical predictions when pursuing scientific investigations –** for example, observes and identifies patterns in nature and predicts what happens next (e.g., if told the sky became dark and cloudy, predicts that it will rain; or predicts if a new object will float or sink)	3.3	18.0	3.8	5.2%
K_5. **Communicates scientific information –** for example, records or describes the properties of common objects verbally or through drawings or graphs	3.1	14.0	10.7	7.8%
K_6. **Demonstrates understanding of physical science concepts –** for example, makes observations that different materials have different properties and that objects are made of different types of materials; compares the relative sizes and characteristics of objects; or describes and explains the different way things move	3.2	17.0	7.6	11.8%
K_7. **Demonstrates understanding of life science concepts –** for example, recognizes the five senses and the related body parts; identifies major structures and functions of parts of plants and animals; or	3.3	17.4	5.4	6.1%

describes the similarities and differences in the appearance and behavior of plants and animals				
K_8. **Demonstrates understanding of earth and space science concepts –** for example, identifies that changes in weather occur from day to day and season to season; describes properties of rocks, soil, and water; or identifies that the sun gives light and heat to Earth.	3.2	16.5	7.3	6.1%

Mean item scores were approximately in the middle of the range, although K_1 (3.5) and K_2 (3.6) were slightly higher. The percentage of children who were rated with the highest possible score ranged from 14.0 to 23.7 percent. This is higher than the percentage of children who were rated "proficient" in the ECLS-K Grade 3 Science ARS scores which ranged from 8.5 to 12.7 percent across those items. Although items K_1 (23.7%) and K_2 (23.4%) had relatively higher percentages of children who were scored as "proficient," they still showed strong correlations with the total score (rs = .84 and .87, respectively). These correlation coefficients are similar in magnitude to those found with the other Science ARS items (rs of other items ranged from .84 to .88).

The percentage of children who were rated with the lowest possible score ranged from 2.6 to 10.7. In comparison, the percentage of children who were rated "not yet" in the ECLS-K Grade 3 Science ARS scores which ranged from 2.7 to 5.4 percent across those items.
Internal-consistency reliability was very good for the kindergarten Science ARS items. Cronbach's coefficient alpha was calculated to examine internal-consistency reliability. The alpha coefficient was α = 0.96. "Not applicable" was recoded to missing in these reliability analyses. This coefficient is similar in magnitude to the alpha coefficients obtained for the Grade 3 Science ARS (α = .95) and the various kindergarten ARS (α all above 0.90). Deletion of items K_1 or K_2 does not change the magnitude of the alpha coefficient.

Average Science ARS scale scores significantly increased with achievement levels of rated kindergarten children. Scale scores were calculated by taking the arithmetic mean of the eight items ("not applicable" was recoded to missing prior to calculating scale scores). A one-way analysis of variance (ANOVA) found that scores for rated children differed by performance level (F (2, 420) = 206.9; p < .0001). The highest achieving students had the highest scale scores 4.4. The lowest achieving students had the lowest scale scores (2.1). Children with average achievement had scores in the middle (3.3). Post-hoc Scheffé tests found that all three scores significantly differed from each other (p < .05).

First Grade

Of the 423 Science ARS forms completed by first grade teachers, valid achievement level data were obtained for 415 forms. Eighty-three forms (20.0%) were for children who were the teachers' highest achieving student; 91 forms (21.9%) were for children who were the teachers' lowest achieving student; and 241 (58.1%) were for children with average achievement.

Means and variances of first grade Science ARS item data showed good variation. For each item, the full range of possible item values (i.e., 1-5) were used by first grade teachers. As shown in table 3, mean item scores were approximately in the middle of the range. The percentages of children who were rated with the highest possible score ranged from 6.7 to 19.7. This range overlaps the range of percentages of children who were rated "proficient" across the ECLS-K Grade 3 Science ARS items which ranged from 8.5 to 12.7 percent. Although F_1 had a relatively higher percentage of children who were rated as "proficient," it still showed strong correlations with the total score ($r = .80$). It should be noted that this correlation coefficient is smaller in magnitude to those found with the other Science ARS items (rs of other items ranged from .84 to .88).

The percentages of children who were rated with the lowest possible score is acceptable, ranging from 3.6 to 10.0. . In comparison, the percentage of children who were rated "not yet" in the ECLS-K Grade 3 Science ARS scores which ranged from 2.7 to 5.4 percent across those items. Two items had relatively higher percentages of children who had not yet had the respective skill taught at the time of the rating: F_6 (46.1%) and F_8 (33.3%).

Table 3. Average item scores across the completed first grade Science ARS forms and percent where the highest possible rating, the lowest possible rating, and "Not Applicable Or Skill Not Yet Taught" was selected.

	Average Item Score	Percent Highest Possible Rating (5)	Percent Lowest Possible Rating (1)	Percent "Not Applicable Or Skill Not Yet Taught"
F_1. **Uses his/her senses to explore and observe –** for example, moves objects and describes how a push or pull can change the way an object is moving; or observes that some living things closely resemble their parents; or observes and describes properties of rocks, soil, and water; or uses tools (such as hand lenses, thermometers, rulers) to gather information about objects around them.	3.4	19.7%	3.6%	2.6%
F_2. **Forms explanations based on observations and explorations –** for example, explains the best growing conditions for a plant after investigating with light and water; or concludes that earthworms come out of the soil because it's raining after paying attention to the sidewalks on a rainy day.	3.1	12.6	8.1	7.6
F_3. **Classifies and compares living and non-living things in different ways –** for example, classifies vegetables that grow above or below the ground; classifies different sounds as either low pitch or high pitch; or measures objects and classifies them by size or weight.	3.3	14.5	6.2	14.7
F_4. **Makes logical predictions when pursuing scientific investigations –** for example, predicts whether or not objects are magnetic based on the materials they are made of.	2.9	8.1	10.0	20.4
F_5. **Communicates scientific information –** for example, records data from measurement tools (e.g., clocks, thermometers, etc.) or constructs bar graphs.	3.0	11.4	7.1	5.9
F_6. **Demonstrates understanding of physical science concepts –** for example, identifies the three states of matter; identifies that heat causes change and compares objects according	3.0	6.7	8.3	46.1

to temperature; or compares the way different objects move (in straight line, by vibration, in a circle).				
F_7. **Demonstrates understanding of life science concepts –** for example, understands that living organisms inhabit various environments and have various external features to help them satisfy their needs; differentiates between those living things that closely resemble their parents (e.g., chick) and those living things that do not (e.g., tadpole); or recognizes that all plants and animals have basic life needs (e.g., air, water, food, etc.).	3.4	15.4	4.3	13.5
F_8. **Demonstrates understanding of earth and space science concepts –** for example, describes how weather affects people's daily activities, describes how land and water store heat from the sun and then warm the air over the land and water; explains that shadows are caused when sunlight is blocked by objects; or identifies natural resources.	3.1	10.5	8.1	33.3

Internal-consistency reliability was very good for the first grade Science ARS items. An alpha coefficient of $\alpha = 0.96$ indicated that there was very good internal-consistency reliability in the first grade data, similar to those found in ARS data from the ECLS-K. Deletion of item F_1 does not change the magnitude of the alpha coefficient.

Average Science ARS scale scores significantly increased with achievement levels of rated first grade children. Similar to the analyses with kindergarten children, first grade scale scores were calculated by taking the arithmetic mean of the eight items ("not applicable" was recoded to missing prior to calculating scale scores). A one-way ANOVA found that scores for rated first grade children differed by performance level ($F_{(2, 414)} = 166.5$; $p < .0001$). The highest achieving students had the highest scale scores 4.1. The lowest achieving students had the lowest scale scores (2.0). Children with average achievement had scores in the middle (3.2). Post-hoc Scheffé tests found that all three scores significantly differed from each other ($p < .05$).

Second Grade

Of the 390 Science ARS forms completed by second grade teachers, valid achievement level data were obtained for 384 forms. Eighty forms (20.8%)

were for children who were the teachers' highest achieving student; 80 forms (20.8%) were for children who were the teachers' lowest achieving student; and 224 (58.3%) were for children with average achievement.

Means and variances of second grade Science ARS item data showed good variation. For each item, the full range of possible item values (i.e., 1-5) were used by second grade teachers. Mean item scores were approximately in the middle of the range (table 4). The percentage of children who were rated with the highest possible score ranged from 3.9 to 26.1 percent. S_1 had 26.1 percent of children as "proficient," but still showed strong correlations with the total score ($r = .70$), although this correlation coefficient is smaller in magnitude to those found with the other Science ARS items (rs of other items ranged from .83 to .88).

The percentages of children who were rated with the lowest possible score is acceptable, ranging from 2.6 to 9.3. In comparison, the percentage of children who were rated "not yet" in the ECLS-K Grade 3 Science ARS scores which ranged from 2.7 to 5.4 percent across those items.

Two items had relatively higher percentages of second grade children who had not yet had the respective skill taught at the time of the rating: S_6 (47.8%) and S_8 (33.1%).

Table 4.   Average item scores across the completed second grade Science ARS forms and percent where the highest possible rating, the lowest possible rating, and "Not Applicable Or Skill Not Yet Taught" was selected.

	Average Item Score	Percent Highest Possible Rating (5)	Percent Lowest Possible Rating (1)	Percent "Not Applicable Or Skill Not Yet Taught"
S_1. Uses his/her senses to explore and observe – for example, compares and classifies objects according to two or more physical attributes (e.g., a basketball is round and has a rough texture, a feather is soft and is 7 centimeters long); or uses observations through the senses to predict an outcome of a simple investigation such as a marble will roll with a greater speed if a ramp is raised 2 cm.	3.6	26.1%	2.6%	1.8%
S_2. Forms explanations based on observations and explorations – for example, explains why one boat floats and another does not; or concludes that a candle stays lit longer under a larger jar because there is more oxygen available; or explains how many layers of clothing provide insulation against heat loss.	3.1	12.9	7.2	6.2
S_3. Classifies and compares living and non-living things in different ways – for example, compares living things based on life cycle; or compares mixtures based on size and/or substance; or describes differences in how the environment affects living things (e.g., migration of birds as the availability of food becomes less when autumn changes to winter) versus how it affects non-living things (e.g., erosion of rocks, evaporation of water).	3.3	12.9	3.9	10.3
S_4. Makes logical predictions when pursuing scientific investigations – for example, predicts the outcome of a simple investigation and compares result with prediction, such as predicting if a plant will grow best in direct sunlight or in shade.	3.3	16.7	4.6	4.4
S_5. Communicates scientific information – for example, records data gathered using simple equipment in	3.2	12.9	6.2	10.3

simple investigations (e.g., changes in weather conditions); summarizes data using charts or graphs; or uses correct units of measurement when recording or summarizing data.				
S_6. **Demonstrates understanding of physical science concepts –** for example, describes the effects of electrically charged materials and magnets; or explains that sound is made by vibrating objects and describe its pitch and loudness.	2.7	3.9	8.2	47.8
S_7. **Demonstrates understanding of life science concepts –** for example, explains that the sequential stages of life cycles are different for different animals, describes how living organisms depend on each other and their environments for survival; identifies differences between living and nonliving objects; or describes how the environment influences some characteristics of living organisms.	3.2	12.6	4.6	22.9
S_8. **Demonstrates understanding of earth and space science concepts –** for example, describes the effects of weathering and erosion; the relationship between the Sun and the Earth; the use of tools to measure weather conditions; or the processes involved with soil formation.	2.9	8.3	9.3	33.1

Internal-consistency reliability was very good for the second grade Science ARS items. Internal-consistency reliability of the second grade Science ARS data was very good with an alpha coefficient α = 0.95, similar to those in the ECLS-K ARS data. Deletion of item S_1 does not change the magnitude of the alpha coefficient.

Average Science ARS scale scores significantly increased with achievement levels of rated second grade children. A one-way ANOVA found that average scores for rated second grade children differed by performance level ($F_{(2, 382)} = 194.8$; $p < .0001$). The highest achieving students had the highest scale scores (4.2). The lowest achieving students had the lowest scale scores (2.0). Children with average achievement had scores in the middle (3.3). Post-hoc Scheffé tests found that all three scores significantly differed from each other ($p < .05$).

Summary and Recommendations

Generally, average item scores were in the middle of the possible range and showed good variation with teachers using the entire range of scores. Some items, most notably K_1, F_1, and S_1 had higher percentages of children rated as "proficient," although they showed strong correlations with the total score. Their deletions from their respective scales do not impact the internal consistency of the scale.

Two items in first grade and second grade tapped science skills or knowledge (Demonstrates understanding of physical science concepts and Demonstrates understanding of earth and space science concepts) that had higher frequencies of not being taught yet to the rated student, which may not be surprising in the beginning of the school year. Data across all three grades showed very strong internal consistency reliability, as well as expected differences across achievement levels of students.

We recommend fielding the kindergarten, first grade, and second grade Science ARS instruments in the ECLS-K:2011. In their review of the instruments (prior to this report) NCES has requested that the number of examples listed under each item be shortened to two examples to ease the burden on respondents. We will make these edits as requested.