Assessing informal learning in data literacy

Daniel Zalles
Judith Fusco
Anna Werner
Reina Fujii

SRI International
333 Ravenswood Ave.
Menlo Park, CA

TERC Inc. developed Kids' Survey Network (KSN), an apprenticeship network. The goal was to prepare the nation's future adults to understand and use survey data that are pertinent to everyday concerns and public opinions of government, business, science, health, and other domains affecting the quality of our everyday lives. Working in teams, KSN participants, ages 11-14, developed their research, survey question writing, and data analysis capabilities through playing games, taking surveys, watching videos, providing feedback to other KSN teams, engaging in challenging online learning activities, and completing their own survey projects on topics of their choice. The project's research study examined the apprenticeship network as a model for delivering challenging activities linked to STEM learning in youth informal learning settings.[1] The project was funded by the National Science Foundation (NSF DRL 0741601).

SRI International evaluated the project. The evaluation team was led by Dr. Daniel R. Zalles, Senior Learning Scientist in SRI's Center for Technology and Learning. This paper describes one component of the evaluation, the assessment of cognitive impacts through collaborative youth group performance on pre-post short answer constructed response tasks that were referred to as "brainteasers". Following up on recommendations by the KSN advisory panel, these "brainteaser" tasks were designed to allow the KSN evaluators to identify gains that groups of youth achieved in KSN activities at pilot implementation sites. The pilot sites were out-of-school time (OST) programs in Oakland and East Palo Alto, California. The brainteaser tasks aligned to the major learning objectives that were addressed in KSN's components (i.e., data games, youth survey projects, and videos of youth talking about doing surveys). SRI designed four sets of brainteaser questions in a progression of increased cognitive challenge. Each of the four sets of tasks addressed a different learning construct:

- bias in survey item wording in relation to a survey purpose.
- appropriate plot structures for representing survey data
- interpretability of data in relation to the categorization scheme
- interpretability of data in relation to graph structure

Each set of brainteaser tasks contained a master prompt and four tasks, and each task required a response to a question. As the participants worked through the increasingly challenging tasks within a set, they need to use increasingly sophisticated thinking. Appendix A contains the complete protocol of brainteaser tasks, plus commentary (in italics) about the purpose of each task and how it fit into a progression. Appendix B contains a detailed group of constructed response assessment results from Group A at a KSN pilot in the OST site in Oakland CA, which was operated by Citizen Schools.

[1] https://www.terc.edu/work/1107.html

The intention was that the tasks be administered collaboratively to the small groups of youth who worked together on their KSN survey projects. However, due to attrition and attendance problems, this intention could not be fully executed. Hence the results provide less of a comprehensive connection between learning and group advancement, though they do provide examples of real learning on the part of various individual youth who expressed responses to the tasks. In other words, the results provide evidence about whether learning gains were occurring at least among some of the youth, although it is not appropriate to try to glean from these results how many of the youth per group made progress. This usually meant that one or two of the youth responded and then the administrator asked the others if they agree. Some youth were more participatory than others.

Another intention was that the tasks be delivered multiple times during the implementation so that there would be a trail of evidence that could be aligned to different group survey project milestones. However, this strategy proved to not be feasible because there were so many constraints on KSN time that it was deemed inappropriate to sacrifice any more time to these tasks beyond once at the beginning and once at the end.

The brainteaser tasks were administered to small groups of youth rather than to individuals because the project team decided that there needed to be a different form of outcome data to triangulate with pre-post-multiple choice assessments, which were administered to youth as individuals. In addition, there was an assumption that, since KSN was designed to have youth work in groups and advance together rather than as individuals, it would be appropriate to look at how much the youth were learning collectively. Unfortunately, due to a combination of student absences and attrition on the two brainteaser administration days, the groups did not maintain 100% cohesion from pre-to post. Nevertheless, due to the fact that all individual youth were tracked and participated in the pre-and post-administrations, even when the group compositions were not identical at both administrations,[2] it was possible to use the results of the brainteasers for answering this question: To what extent did the brainteaser results yield evidence that intended learnings were occurring among at least some of the youth?

Scoring procedures were as follows. Per task, if a particular group failed at pre-time or got a partial pass, the brainteaser administrator asked the group again at the post. If they passed at pre, they were not asked again at post. The order of presentation of the tasks was prioritized so that all on which the group got "F's" (for failure) at the pre received the highest priority for attention at the post, and partial passes at the pre had secondary priority. Unfortunately, time ran out before an opportunity arose to cover all of the PP tasks again. (It needs to be noted that this terminology about whether the youth "passed" or "failed" a task was not communicated to them. They were not told how they performed except at the post, and only when they arrived at a correct solution.)

The brainteaser instrument was revised somewhat based on how youth responded who were participating in a prior piloting at a summer camp program in East Palo Alto, CA. It was determined that some of the tasks needed to be re-worded to have clearer intent, one needed to be added, and two needed to be eliminated. In addition, certain plots needed to be somewhat restructured to eliminate elements that could lead youth to misinterpret structures in a way that would distract them from identifying an appropriate solution to the problem posed in the task.

[2] One girl in Group A at the Citizen School pre-administration of the brainteaser tasks was reassigned to Group B at the post administration because the initial third person from Group B left the program in the meantime due to a family issue. Hence, this reassignment was a way to maintain three kids per group rather than 2 kids in one group and four in the other

Hence, the instrument used with the East Palo Alto youth was somewhat different from the one used subsequently with the Oakland youth yet in both cases, the pre-version was parallel to the post.

Appendix A

Brainteaser tasks

TASK SET A. QUESTION CONSTRUCTION

1. Lionel wants to know how happy kids are wearing uniforms to school so he asks on a survey:

How happy are you with wearing uniforms to school and having a longer school day?
a. Very happy
b. Kind of happy
c. Not happy

What do you think of his question and answer choices? Will they give him the information he's looking for? If you don't think so, suggest improvements.

(This is a double barreled question. The question should only be about wearing uniforms. Kids who don't feel the same way about uniforms and longer school day would not be able to answer honestly)

2. Alicia wants to know how much kids like fruit and how much they like vegetables so she asks on a survey:

Do you like fruit or do you like vegetables?
a. Yes
b. No
c. Not sure

What do you think of her question and answer choices? Will they give her the information she's looking for? If you don't think so, suggest improvements.

(This question forces respondents to choose whether they like fruits or vegetables. It should be rewritten into two separate questions, one about fruit and the other about vegetables, or the questioned choices should be fruit, vegetables, both, and neither)

3. Carol wants to know if kids at his school think that the principal should allow kids to arrive at school 10 minutes late every morning without getting in trouble.

Do you feel that the principal should allow kids to arrive at school 10 minutes late every morning without getting in trouble?
a. Yes, very strongly
b. Yes, but not very strongly

c. Not sure

What do you think of her question and answer choices? Will they give her the information she's looking for? If you don't think so, suggest improvements.

(The question is fine but the selection choices do not allow for somebody to express they don't want it. Adding a "No" choice would be sufficient)

4. Charlie wants to know if kids at his school like playing video games, so he asks on a survey:

"How often do you play video games?"
a. "At least once a month"
b. "At least once a week"
c. "Every day"

What do you think of his question and answer choices? Will they give him the information he's looking for? If you don't think so, suggest improvements.

(The configuration of question and selection choices is biased because the question promotes a positive answer and the answer choices only allow for a positive answer)

Legitimate improvement:
Student could suggest that the question be reworded to something like
Do you like playing video games?
a. A lot
b. A little
c. Not at all

5. Bob wants to know how much time kids spend watching TV each week, so he asks on a survey:

"How much work time do you spend watching TV each week?
a. A lot
b. A little
c. Not much
d. None at all

What do you think of his question and answer choices? Will they give him the information he's looking for? If you don't think so, suggest improvements.

(The choices are too ambiguous and will mean different things to different respondents. There need to be more exact choices specifying quantitative ranges of values such as more than 10 hours, 1 to 9 hours, etc.)

TASK SET B. APPROPRIATE PLOT STRUCTURES FOR REPRESENTING SURVEY DATA IN WAYS THAT PERMIT CONDUCTING ANALYSES NEEDED TO FULFILL SURVEY PURPOSES

1. George did a survey to see what kinds of animals kids prefer as pets. Kathy, Cindy, and John said they prefer dogs. Roger, Bob, Annie and Sally said they prefer cats. Emma and Linda said they prefer rabbits, and Ron said he preferred snakes. What is the best way to plot these results? Sketch a plot.

(This task requires that the youth convert text-displayed results to results displayed on a data plot. It requires understanding simple stack plot structures).

2. Marie conducted a survey to see what kind of snack foods kids like. She thinks that kids like cookies the most. She put the results of her survey in this table:

	Cookies	**Fruit**	**Ice Cream**	**Cake**	**Chips**
Sandy	1				
Cody		1		1	
Terry	1				
Clarence			1		
Jeremy		1			1
Barbara			1		
Rick	1				
Amy				1	
Jackie			1		
Henry	1				
Francis					
Ian			1		
Mark					
Simon			1		

What is the best way to plot these results so that Marie can quickly check if she's right? Sketch your plot.

(This task requires that the youth convert table-displayed results to results displayed on a data plot. As with level I, this requires that they understand simple stack plot structures but this time they also need to understand the table structure shown in the prompt rather than just text.)

3. Tom conducted a survey to see if men are more likely than women to want the government to lower taxes. He thinks they are.

Name	**Gender**	**Attitude about lowering taxes**

Henry	man	against
Sue	woman	for
Sam	man	for
Alice	woman	for
Sally	woman	against
Zoe	woman	for
Ralph	man	against
Homer	man	against
Mike	man	for
Jasmine	woman	against

What is the best way to plot these results so that Tom can see if he's right? Sketch your plot.

(This task requires that the youth convert table-displayed results to results displayed on an axis plot. As with level 2, this requires an understanding of table structures this time around, the more complex axis plot structures rather than the simple stack plot structures of levels 1 and 2. It would be sufficient if the youth' plot shows political identity as one axis and attitude about healthcare on the other axis.)

4. Sharon thinks that people from California are more in favor of lowering taxes than people in New York, regardless of whether they are men or women. To check his hunch he surveyed some adults. The table below shows her results.

Name	**Gender**	**State**	**Attitude about lowering taxes**
Liz	woman	California	for
Frank	man	New York	against
Simon	man	California	for
Laura	woman	New York	against
Dan	man	New York	for
Candace	woman	California	for
Marcy	woman	New York	for
Marshall	man	New York	against
Judy	woman	California	for
Josh	man	California	against

What is the best way to plot these results so that Sharon can check her hunch. Sketch your plot.

(This task requires that the youth convert table-displayed results to results on an axis plot where each data point is coded to represent the values of a third variable. As with levels 2 and 3, this requires that the youth understand table structures. As with level 3, it also requires understanding of axis plot structures. However this task brings in a third variable that can be represented by different icon symbols. It doesn't matter which variable is the X, which is the Y, and which is represented by icons on the plot.)

Appendix B

Youth brainteaser results from Group A in Oakland

These findings cover the results from one of the two groups carrying out the "brainteaser" tasks. Unfortunately, the second group had too many disruptions from certain poorly behaving youth to get very far. An additional problem for both groups was that there was not much time allotted by the OST program for this activity. The youth had to be pulled from the cafeteria and given these tasks during their short lunch period. About 2/3rds of the way through the tasks, the youth had to be moved from where the SRI researchers started the administration (which was in the same room where the OST program staff were having a loud meeting and where other youth kept coming in and out) to the library where it was much quieter. The SRI researchers had 25 minutes with the youth which included a few minutes to pass between rooms and get resettled, and also to pass out lunch. They were eating their lunches while responding to the tasks.

The first group showed progress from pre-to post and those results are reported below. The results for that group provide evidence that learning was taking place among some of the youth, although it is not appropriate to try to determine from these results how many of the youth specifically made progress. Usually this meant one or two of the youth responded and then the administrator asked the others if they agree. Some youth were more participatory than others.

KEY:
P=Passed
PP=Partially passed
F=Failed

Per task, if the youth group failed at pre-time or got a partial pass, the intention was to ask them again at the post. If they passed at pre, they were not asked again at post. The order of presentation of the tasks was prioritized so that all the tasks on which the group got "F's" at the pre received the highest priority for attention at the post, followed by the partial passes from the pre. Unfortunately, time ran out before an opportunity arose to cover all of the PP tasks again at the post.

All tasks where growth was shown from pre-to post are denoted with an asterisk. On no tasks did youth who received partial passes on the pre receive failures on the post.

STUDY A:
Broadly speaking, the youth went from critiquing questions in response options based on their reasonableness with regards to the topic to critiquing them based on their construction as valid survey tasks. This was a sign that they were beginning to understand aspects of survey research methodology, which is an intention of the program.

Task 1A. *

- Pre: Group thinks it's a good question and does not explain why (F)
- Post: According to the group, the problem is that "It's asking you 2 questions at once" (P)

Task 2A. *

- Pre: Group thinks it's a bad question because "they might not all be honest... because if they say that they like vegetables, then the school board could take away the fruit (F).
- Post: Group recognized you need to split into two separate questions-- suggested change to "which do you like, fruit or vegetables? Or how much do you like vegetables or how much do you dislike fruit (P)

Task 3A. *

- Pre: Group thinks it's a good question again because of the content, then talked about something regarding if you're late, you have time to eat breakfast and if you're expelled from being late too much, you can't catch up. (was hard to decipher what he said) (F)
- Post: Group recommended to change the answer choices to yes, no, or not sure (Though some were focusing on the appropriateness of the content of the question rather than methodological issues about validity and reliability of the question) (P)

Task 4A. *

- Pre: Group said it's a bad question because it's "too short". How often do you play video games every day? Or, do you play every day, or so you now play every day) (PP)
- Post: Group properly recognized that the question doesn't ask them whether they like video games but how often they play them there should be more choices - (that comment is either off-track or not developed enough) (P)

Task 5A.

- Pre: Group thinks it's a good question. (F)
- Post: Group said "everybody watches TV (changed answer choices to yes, no sometimes, not sure) (F)

STUDY B

Task 1B.

- Pre: The administrating SRI researcher asked the group what a plot is. The group provided the definition of plot as a story element, not of a data plot. Yet, responding to the administrating researcher encouraging them to think about it all that more, one youth made the association with matrices. He said he did some "plots" before, in fifth grade, where he learned about how one can put on a table different icon symbols, so he drew two tables -- one with "1"'s for the information and one with "tally marks" (that was his language). He also then described how one could use colors as well; for example orange if they prefer dogs, blue if they prefer cats, etc. He drew some tables which properly show rows and columns, where the rows correctly show each individual kid and the columns show the types of animals the kids prefer. The cells indicate the intersection between preferred pet and student respondent. (PP)
- Post: Group says count the people to show how many people voted on the same animal (PP)

Task 2B.

- Pre: Student made a sufficient plot -- a bar graph where each bar was labeled a different type of food and the y axis showed 1 to 6 (P)
- Post: The administrating SRI researcher didn't ask because they passed this at pre-time

Task 3B.

- Pre: A student plotted. His plot summarizes the data (a small table showing the number five and that the five means women). Hence, he grasped only that there were five men and five women but he didn't show on his plot how many people wanted to lower taxes.

The researcher's interpretation is that the student knew that he wanted the plot to be useful for checking the hunch about taxes (which was the purpose of the plot as indicated in the question), but he didn't understand how to do that. What was needed was an XY plot. (PP)

- Post: The group suggested a stack plot with "women" and "men" on the X. axis and numbers from 1 to 10 on the Y. axis. So they were halfway there. Unfortunately, they did not recognize where to put the data showing attitude about lowering taxes. (PP)

Task 4B.

- Pre: They did not try to do this one. One student said "it's too difficult." (F)
- Post: The administrating SRI researcher did not get to ask this one at the post because it became low priority due to the assumption that if they didn't get task 3 they wouldn't get this one correct either, plus time ran out.

STUDY C

Task 1A. *

- Pre: They could not get off the ground on this task. When the researcher asked if they thought these were good choices, a student said "it would have been better if they (the kids in the task) had done a plot". This comment shows that he is trying to say something intelligent but what he said is unrelated to the question. Then the researcher read each choice, and somebody said "playing sports". The researcher responded by saying, "what if you could only give one answer, would those still be good choices?" The youth said yes. They were not picking up at all on the problem, which is that these choices were not mutually exclusive. (F)
- Post: Group said to get rid of watching comedy shows and playing basketball. One comment made was that "watching television and watching comedy shows is the same thing" (P)

Task 2A. *

- Pre: The group could not recognize that the choices are bad because they are too subjective. What they focused on instead was whether they would personally prefer the different types of movies listed. Student A said he thinks some kids would want good soundtracks because they like music. Student B replied "I don't like music." Student A then said, "I think half of the people won't come because they don't like music." Student B then replied about how movies with good actors are preferable because then you can understand the movie better, or something to that effect (the audio was difficult to discern at this point) (F)
- Post: The group said to change the choices to genres ("action, comedy, thriller, drama") Note: they did not say genres per se, but that is what they meant (P)

STUDY D

Task 1D.

- Pre: To see if the group could understand the basic plot structure, he asked them to explain the very first icon on the upper left side ("what does it mean here when you see this little face?") A student responded that "it means how many people?" The researcher then asked who is this person. The student properly answered "Aida." Then the researcher asked her if she likes the plot and she said yes and explained why. From this dialog, one can conclude that the student knew how to interpret what was on the plot and was

satisfied with it because she said that she can identify from the names which of the respondents are boys and which are girls. Hence, she was arguing correctly that the plot was good for checking the driving conjecture but was not perceiving how it would be better to reduce the data on the X. axis to two values: boys and girls. (PP)

- Post: Didn't ask -- ran out of time

Task 2D.

- Pre: When the researcher asked if the plot shows that younger kids are more or less likely to obey their parents, a student said "older" because the older ones "can get kicked out." Hence, his response focused on the content rather than the data structures. The researcher then said, "Yes, I see, but can you tell that from this plot?" Student said yes. He counted out to the number 12 from plot icons and said, "Mostly everybody liked obeying their parents." He kept looking at the plot and seemed to all of a sudden get a little confused but perhaps he was not interpreting the data correctly and he said, "Not always." He then read the top line of the plot (for 12th grade) and correctly interpreted what the icons mean So, in doing so, he also showed he is capable of interpreting what the specific icons mean on an XY plot, Just like the student who was the main responder for Task D1, he could not properly critique the goodness of the plot structure. (PP)
- Post: Didn't ask -- ran out of time

Task 3D. *

- Pre: They couldn't respond to this one (F)
- Post: Student A said she could not tell if the plot is useful. The researcher replied that Isabel (the "girl" in the prompt) thinks that six and seventh graders probably like rock music the most and the question is if the plot shows this. In other words, The researcher was reminding the group to stay focused on sixth and seventh grade in particular. A breakthrough in their thinking came when the researcher rephrased the question: He said, can Isabel tell whether sixth and seventh graders like rock music more than the others do? (As opposed to the stem of the question which said "...like rock music the most"). Then, a different student in the group made a comment which shows some understanding: "It doesn't show which ones are for 6th and which are for 7th" (P)

Task 4D.

- Pre: At first, the group could not respond to this task. The researcher tried to provide some prompting on it. He backed up to ask a basic question to see if the youth understood what the data points were saying in the first place. He asked: "Do you understand what any of these heads stand for?" Student A said that the heads stand for how many seniors there how much they like each subject, then Student A corrected herself and said no, it's about whether they're interested in basketball. She also said that she thought that the table numbers were percents. The researcher pointed to one of the uppermost of the five icons of respondents between 50 and 75 inches tall. About this particular data point/icon, Student B said correctly "He's not really interested much in basketball. (The icon was in the "rarely interested." column). Then, Student B appropriately interpreted the icon as symbolizing that this particular student is about 55 inches tall (though the plot structure is only saying that the student represented by that icon is somewhere between 50 and 75 inches tall, not specifically 55 inches tall. *(PP)

- Post: Didn't ask -- ran out of time