LONG-TERM IMPACT ON STUDENTS' INFORMAL INFERENTIAL REASONING

Einat Gil and Dani Ben-Zvi
University of Haifa, Israel
egil767@gmail.com , dbenzvi@univ.haifa.ac.il

Long-term effects of learning are a desirable outcome of any educational program and are far from being an obvious result in education. Furthermore, statistical concepts tend to be ambiguous and "short lasting" in students' reasoning, even among tertiary students. In this longitudinal study, long-term impact of teaching and learning was sought among ninth graders, three years after their participation in a three-year intervention (grades 4-6) of the Connections Program. In a mixed methods study, students from two groups – those who have / have not taken part in the program – were closely followed and compared throughout three extended data inquiry activities and took a statistical knowledge and thinking proficiency test. Results and implications are presented.

THEORETICAL BACKGROUND

Long-lasting results of educational programs are highly desired and are not taken for granted in statistics education, where even short-term conceptual understanding is challenging. This is specifically so regarding statistical Informal Inferential Reasoning (IIR, Makar, Bakker, & Ben-Zvi, 2011), which requires an integration of three complex components: Generalization, data-based evidence, and uncertainty (Makar & Rubin, 2009). IIR refers to "the cognitive activities involved in informally drawing conclusions from data (samples) about a 'wider universe' (the population), while attending to the strength and limitations of the sampling and the drawn inferences" (Ben-Zvi, Gil & Apel, 2007). While serving as a possible bridge between Exploratory Data Analysis (EDA) and formal statistical inference, various aspects of IIR became a hub of interest in recent studies, such as: Reasoning about sampling variability, the role of context, technology and explanations in developing statistical reasoning, and the design of tasks and learning trajectories (e.g., Biehler, Ben-Zvi, Bakker, & Maker, 2013; Gil & Ben-Zvi, 2011; Makar & Ben-Zvi, 2011). IIR is defined in this paper as including reasoning about sample and sampling as well as reasoning about informal inference.

A learning environment that provides students rich extended opportunities of active learning for understanding of complex key concepts, the interrelations among them, and their application can have a long-term influence on the ways their knowledge is constructed and their ability to preserve and use it (Bransford, Brown & Cocking, 2000). Studies of long term effect of students' reasoning (e.g., two years after an intervention) are relatively rare and typically take the path of evaluating the effects of large scale interventions. They are common in the study of under-privileged population programs, focusing on students' grades, IQ, or special education placement (e.g., Barnett, 1995). These are usually large-scale quantitative studies, but some are small-scale programs that look at shorter term (a couple of months) effects (e.g., Zohar & Ben-David, 2008).

METHOD

The Research Question

To study the long-term impact of the Connections Program, we compared the characteristics of ninth graders' IIR three years after they took part in it, with the IIR of ninth graders from the same school who did not participate in this program. The following research question guided the study: *What are the characteristics of IIR among Connections graduates in comparison to non-Connections students?* This question was examined using mixed methods, i.e., a combination of quantitative and qualitative methods (Tashakkori & Teddlie, 2003), to explore and compare students' IIR through their verbalization and visualization of relevant statistical concepts.

The Setting

The Connections Program included a research- and inquiry-based learning environment for grades 4–6 (2005–2007) that aimed to develop students' statistical reasoning and especially IIR (Ben-Zvi et al., 2007). It took place at a science-focused school in Israel in which most of the

In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, Arizona, USA.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

students were from affluent backgrounds. The program started in grade 4 with a five-week intervention focusing on distribution and EDA, and continued in grade 5 with a five-week intervention focusing on sample, sampling and IIR using the ‘growing samples’ pedagogic heuristic (Bakker, 2004; Ben-Zvi, 2006). In grade 6, the five-week learning trajectory aimed at deepening students’ reasoning about sampling and informal inference. For example, students in grade 6 were asked to explain their informal inferences from random samples and argue about them as they engaged in authentic collaborative data-based inquiries using *TinkerPlots* (Konold & Miller, 2005).

The ninth graders in the current study were Connections graduates (CG) and Non-Connections students (NCS). The NCS joined the school in grade 7 and studied in mixed classes with CG. Within the CG cohort, we distinguished a smaller group, who participated in the focal class of the program (CGF) that received more intervention hours than others. The groups were similar academically and did not study sampling and inference in middle school.

Quantitative Study

117 students (68 CG students, among them 25 CGF students, and 49 NCS) took a knowledge and thinking proficiency test. The test focused mostly on sampling and IIR but also on data analysis, question posing, and research skills. It included ten closed and open questions, some of which had multiple items and a request to provide explanations. The test was partially based on a Connections sixth grade test but with different problem contexts. The response rate was 92%, out of which six questionnaires had to be removed. After creating a coding system, the data were analyzed using *t*- and *nonparametric*-tests to compare between the groups.

Qualitative Study

We chose five CGF and five NCS pairs of students to highlight differences between the groups. They all participated in three inquiry data-based activities (Table 1). In addition, the NCS pairs participated in an introduction to *TinkerPlots* activity, while the CGF students were offered to refresh their *TinkerPlots* knowledge, though only three students attended.

Table 1. Ninth grade activities.

Activity	Context and statistical content
1	a. Plan a survey on a self-chosen subject, design a sampling method to infer about the student population of Haifa (open-ended activity without a computer) b. “Media activity” on elections’ surveys (structured activity without a computer)
Preparatory	A small inquiry EDA activity to introduce *TinkerPlots* (NCS)
2	An IIR investigation on a self-chosen subject using a random sample from school data (structured activity with *TinkerPlots*)
3	An IIR investigation of several random samples, comparing results and increasing sample size (partially open activity with *TinkerPlots*)

In this paper we present a case study on two focal pairs: group 1 – Odi and Asi (CGF) and group 2 – Alon and Segev (NCS) (pseudonyms, males, age 15). They all had high academic abilities and communicative skills. Asi and Segev were gifted students. According to the intensive sampling method (Patton, 2002), they were chosen to provide a rich source of information.

Two episodes from the students’ work on Activity 2 (Table 1) are briefly discussed below. In this Activity, students explored research questions that they had posed using a random sample (n=30) drawn from a large database about school students. In episode 1, Segev and Alon investigated their research questions: 1) *Do you have a pet? If you do, how old are you, and what is your gender?* 2) *Do you have a pet? If you do, is it a dog or a cat?* 3) *Are pet owners more independent?* They conjectured that there would be more dog than cat owners, and that children who own a pet would be more independent than those who do not own a pet. In episode 2, Odi and Asi investigated one question (out of three they had posed): *Is the number of people at home related with owning pets, and how?* They conjectured that there would be an association between the two attributes. While Asi expected a positive association, Odi was not sure.

To examine and compare the students' IIR, these episodes were fully videotaped and transcribed. The analysis of the videotapes was based on interpretive microanalysis (Meira, 1998), a qualitative detailed analysis of the transcripts, taking into account verbal, gestural and symbolic actions within the situations in which they occurred. The goal of the analysis was to trace and infer students' reasoning about sample and sampling and informal inference from a mainly cognitive perspective, taking into account socio-cultural processes of understanding and learning. After the NCS and CGF students' IIR and explanations were identified and compared by the first co-author, the results were triangulated by experienced researchers. Content analysis (Shkedi, 2003) was also used to compare between emergent statistical ideas and concepts at different points of the pairs' work.

RESULTS

Quantitative results

In a *t*-test for independent samples of the final test mark, a significant difference was found between CG (M_1=52.03, Sd_1=11.91, n_1=68) and NCS (M_2=47.77, Sd_2=10.57, n_2=49); $p<0.05$, (115) t=2.00. The difference was even greater in a Mann-Whitney test for independent samples between the final score of the students CGF to NCS (*U*=403.5, $p<0.05$) (Fig. 1).

Differences in the test subjects were found in IIR between CG (M_1=48.01, Sd_1=13.52, n_1=68) and NCS (M_2=43.20, Sd_2=12.19, n_2=49); $p<0.05$, (115) t=2.01, and in question posing (Mann-Whitney test; *U*=525, $p<0.05$). Differences in IIR were found also between CGF and NCS in a Mann-Whitney test (*U*=404, $p<0.05$), in addition to a difference in reasoning about informal inference (*U*=419, $p<0.05$). There were no significant differences in data analysis and reasoning about sample and sampling, although CG and CGF's means were higher than NCS.

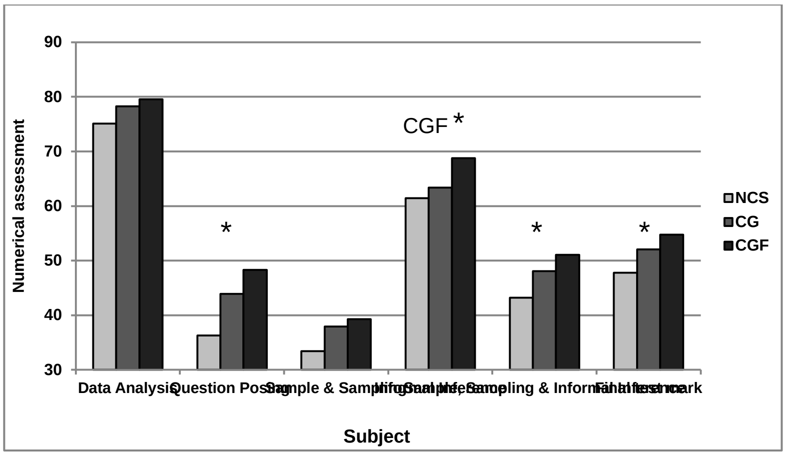

Figure 1. The statistical knowledge and thinking proficiency test results. Significance marked by * is both for CG-NCS and CGF-NCS, unless CGF is written, where it is only for CGF-NCS.

CG and CGF students were also better than NCS in responding to a question on making an inference to an increased sample size by adding points to a graph based on two given graphs of small samples of seventh grader's heights (adapted from Zieffler, Garfield, delMas, & Gould, 2007). In a Mann-Whitney test CGF students' ability to infer to an increased sample size was better than NCS in terms of the spread, center and shape of the distribution (*U*=304.5, $p<0.01$), and CG students were better than NCS in the distribution shape (*U*=1072, $p<0.01$). CG and CGF students' tended to create a normal shape distribution while the NCS tended to create a bi-modal distribution shape (Figures. 2-6).

These visual conceptual findings coupled by other verbal conceptual findings (Gil, forthcoming) provide evidence for a long-term impact of the Connections Program on students' IIR, supported by a qualitative analysis of the two focal pairs, a sample of which is presented as follows.

Qualitative results

Segev and Alon investigated their three research questions with *TinkerPlots*, creating multiple graphs (e.g., Figs. 7-8) and discussed the results. Their conclusions were: 1) *Most of the kids that own pets are girls, most of which are 10–12*; 2) *Children who do not own pets are generally more independent*; and 3) *There are more dog pets than cats.* They also explained that, *most of the kids provided incorrect information about whether they owned a dog or cat,* since they were puzzled by conflicting responses in which some of the children who wrote "no pet" reported owning a dog or a cat.

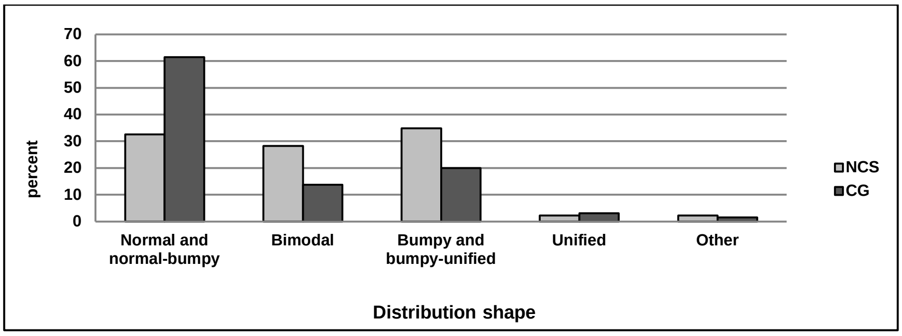

Figure 2. Distribution shape of inference to an increased sample size.

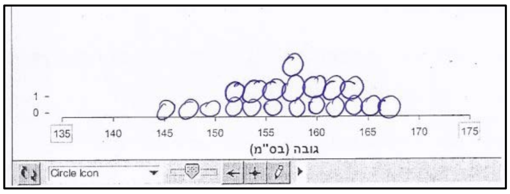

Figure 3. Normal distribution (CG).

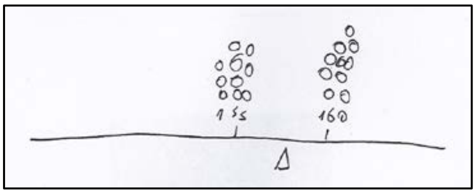

Figure 4. Bi-modal distribution (NCS).

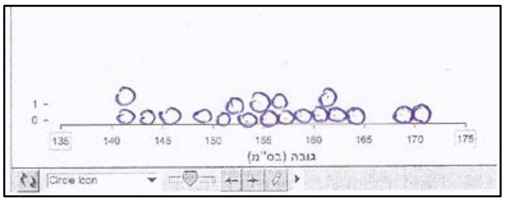

Figure 5. Bumpy distribution (CG).

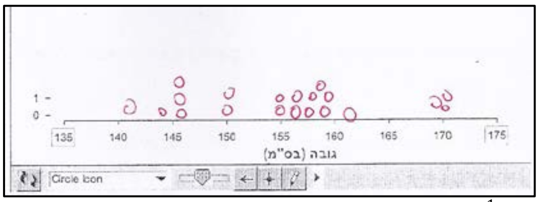

Figure 6. Bi-modal distribution (NCS)[1].

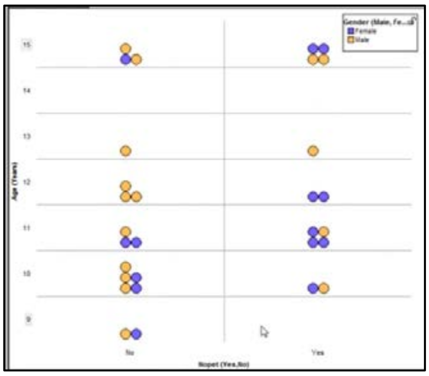

Figure 7. No-pet, age and gender (n=30).

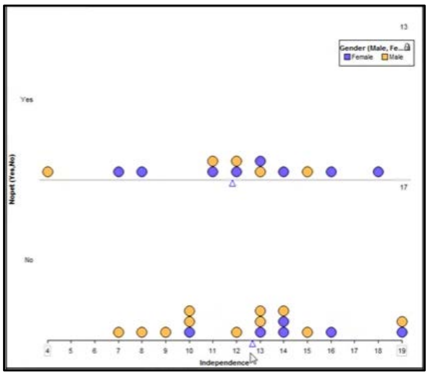

Figure 8. Independence, no-pet and gender (n=30)

Segev and Alon tended to ask descriptive questions about the data, except for the third association question. Their conclusions were mostly interpretations of the sample data rather than informal statistical inferences (e.g. *Most of the kids that have pets are girls*). They did not seem to generalize beyond the sample data, did not provide data-based evidence, and did not express uncertainty explicitly.

Asi and Odi, who investigated one selected research question with *TinkerPlots* (*Is the number of people at home related to owning pets, and how?*), created and worked mainly with one graph (Fig. 9) and discussed the results. Their informal inference was: *Based on the sample, we came to understand that there is no clear trend in the association between the criteria [variables]. We drew lines on both sides [distributions] and it seems that there is no association. The number of items was distributed quite evenly (45%-55%).*

The three questions that Asi and Odi originally posed were all association questions. Their informal inference included the three IIR elements: a data-based generalization in uncertain language (*it seems*, *no clear trend*). It is worth noting that Asi and Odi's informal inferences and their report about the investigation in grade 9 resembled the ways they reasoned in grade 6.

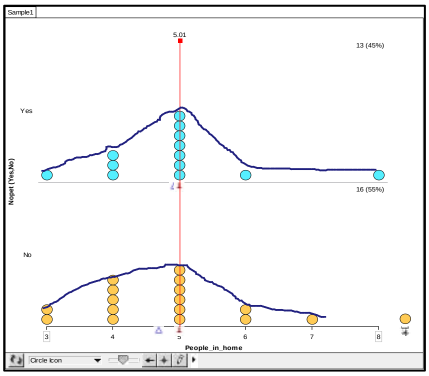

Figure 9. Number of people at home and owning a pet (n=30).

Analysis of these issues among the ten pairs revealed that most of CGF and NCS pairs formulated association rather than descriptive research questions in Activity 2. In this respect, Segev and Alon's questions were not typical among the NCS. However, in relation to IIR, the NCS tended to generalize in a deterministic manner without substantiation in the data at hand or expressions of uncertainty. Thus, the quantitative and qualitative findings suggest that CG students, and particularly CGF, reasoned inferentially significantly better than NCS. These results provide supportive evidence for the long-term impact of the Connections Program on the participating students' IIR.

DISCUSSION

In this paper we presented a snapshot of quantitative and qualitative results from a research on the long-term impact of participation in the three-year Connections Program on ninth graders' IIR three years after its end. The findings suggest the existence of significant differences between the groups' IIR. CG students exceeded NCS in three areas, briefly presented in this paper: 1) conceptual understanding, 2) informal statistical inference, and 3) aggregate view of a distribution.

First, they used statistical inferential concepts in their explanations in a more meaningful and extensive manner. The analysis of students' explanations in the test and the two focal pairs' discourse indicated not only that CG and CGF used statistical concepts more accurately than NCS, but also that they viewed these concepts as a connected conceptual network required for IIR. In this respect, CG and particularly CGF's IIR resembles aspects of experts' statistical thinking (Wild & Pfannkuch, 1999). The second difference, as examined in the ten focal pairs, was that CGF students formulated stronger informal inferences than NCS in terms of the three pillars of informal statistical inference (Makar & Rubin, 2009). This means that CGF were more fluent with respect to IIR, including the uncertainty involved in a generalization from random sample to population, and the need to support their inferences using data-based evidence. Third, CGF students, and in part CG students, were more able than NCS to infer to an enlarged sample, based on the characteristics of the distributions they created. This seems to represent a greater ability to understand a distribution in terms of an aggregate view (Konold et al., in press).

These findings are admittedly limited by the study's idiosyncratic circumstances. We are therefore careful in making clear causal explanations, but CG's advantage in IIR seems to relate to their experiences in the Connections Program: data investigations with *TinkerPlots*, reasoning about data and distributions, drawing informal inferences, and growing samples and resampling. We are well aware that significant challenges are still ahead of us. These include questions such as:

what components of the Connections' design or pedagogy had a greater impact? What are the IIR differences among individual students? Will the same impact occur after relatively shorter programs (e.g., one or two years)?

The evidence we provided on long-term impact of learning statistics on students' statistical understanding and reasoning is encouraging and challenging. We call for more longitudinal studies in statistics education that are crucial for the advancement of the understanding of the long-term effects of our ongoing educational efforts.

ENDNOTE

[1] A "hill" in normal or bi-modal distribution was coded if three dots or more were stacked.

REFERENCES

Bakker, A. (2004). *Design research in statistics education: On symbolizing and computer tools.* Doctoral dissertation, Utrecht University, Netherlands.

Barnett, W. S. (1995). Long-term effects of early childhood programs on cognitive and school outcomes. *The Future of Children*, *5*(3), 25-50.

Ben-Zvi, D. (2006). Scaffolding students' informal inference and argumentation. In A. Rossman and B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute.

Ben-Zvi, D., Gil, E., & Apel, N. (2007). *What is hidden beyond the data? Helping young students to reason and argue about some wider universe.* The Fifth International Research Forum on Statistical Reasoning, Thinking, and Literacy (SRTL-5), University of Warwick, UK.

Biehler, R., Ben-Zvi, D., Bakker, A., & Maker, K. (2013). Technology for enhancing statistical reasoning at the school level. In M. A. Clements, A. Bishop, C. Keitel, J. Kilpatrick, and F. Leung (Eds.), *Third international handbook of mathematics education* (pp. 643-690). Springer.

Bransford, J., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school*. Washington. DC: National Academy Press.

Gil, E. (Forthcoming). *Long term impact of promoting informal inferential statistical reasoning in an inquiry-based learning environment*. Unpublished doctoral thesis, Haifa University, Israel.

Gil, E., & Ben-Zvi, D. (2011). Explanations and context in the emergence of students' informal inferential reasoning. *Mathematical Thinking and Learning*, *13*(1-2), 87-108.

Konold, C., Higgins, T., Russell, S. R., & Khalil, K. (in press). Reasoning about data. *Educational Studies in Mathematics*.

Konold, C., & Miller, C. (2005). *TinkerPlots: Dynamic data exploration, statistics software for middle school curricula.* Emeryville, CA: Key Curriculum Press.

Makar, K., Bakker, A., & Ben-Zvi, D. (2011). The reasoning behind informal statistical inference. *Mathematical Thinking and Learning*, *13*(1-2), 152-173.

Makar, K., & Ben-Zvi, D. (2011). The role of context in developing reasoning about informal statistical inference. *Mathematical Thinking and Learning, 13*(1-2), 1-4.

Makar, K., & Rubin, A. (2009). A framework for thinking about informal inference. *Statistics Education Research Journal*, *8*(1), 82-105.

Meira, L. (1998). Making sense of instructional devices: The emergence of transparency in mathematical activity. *Journal for Research in Mathematics Education*, *29*(2), 121-142.

Patton, H. J. (2002). *Qualitative research & evaluation methods*. Thousand Oaks, CA: Sage.

Shkedi, A. (2003). *Words of meaning: Qualitative research–theory and practice* (in Hebrew). Tel-Aviv University: Ramot.

Tashakkori, A., & Teddlie, C. (2003). *Handbook of mixed methods in social & behavioral research*. Thousand Oaks, CA: Sage Publications.

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, *67*(3), 223-265.

Zieffler, A. S., Garfield, J., delMas, R., & Gould, R. (2007). *Studying the development of college students' informal reasoning about statistical inference*. The Fifth International Research Forum on Statistical Reasoning, Thinking and Literacy (SRTL-5), University of Warwick, UK.

Zohar, A., & David, A. B. (2008). Explicit teaching of meta-strategic knowledge in authentic classroom situations. *Metacognition and Learning*, *3*(1), 59-82.