MIDDLE-SCHOOLERS PRIMED TO REASON COUNTERFACTUALLY ASK MORE INTERESTING QUESTIONS

Sneha Chakravarty[1], Anveshna Srivastava[2*], Koumudi Patil[3],
Poddar International School[1], IIT Kanpur[2,3]
anveshna.sriv@gmail.com[2]

Counterfactual reasoning is a crucial component of scientific inquiry, giving an investigator the ability to design numerous (thought) experiments given a phenomenon. It affords learners the ability to understand complex chain(s) of causal attributions by letting them think about situations where existent factual event(s) are countered in a systematic manner, either by reference to existing data, or by designing new experiments. In this work, we assessed whether asking 5th and 6th grade students (10-11 years old) to reason counterfactually results in a measurable difference in the way they pose questions about a scientific phenomenon - 'biological adaptation'. Our empirical results indicate that the intervention does make a significant difference in the nature of questions asked. Our results have implications for inquiry-based learning, emphasizing the deployment of counterfactual reasoning in science curricula.

INTRODUCTION

Any natural system, be it an organism, an ecosystem, or even an inorganic crystal, arrives at its particular nature through the interaction of multiple factors. Trying to identify how the system would be different if any one factor was changed, *ceteris paribus,* is an excellent way of trying to understand the influence of that factor on the overall system (Minner, Levy & Century, 2002). Thus, the teaching of science offers fertile ground for the application of counterfactual reasoning.

But what exactly is counterfactual reasoning? It is the ability to reason by considering alternatives to the existing fact, or in other words, it is thinking with a 'what if' (Roese, 1997). To illustrate, in an illuminating developmental study, Rafetseder, Cristi-Vargas and Perner (2010) told both adults and 6 year old children a story. A mother could place candy on either the top or the bottom shelf of a cabinet. If she places it on the top shelf, her tall son can reach the candy, but he can't bend down to reach it if it is on the bottom shelf, because he's recently had a fracture and his leg is in a cast. Her small daughter can only reach the candy if it is placed on the bottom shelf but not on the top one. When the researchers asked adults what would happen if the candy was placed on the top shelf and the girl came into the kitchen, they were able to answer correctly 100% of the time. However, only 24% of 6 year olds could answer such counterfactual questions correctly.

This finding exemplifies a view commonly held in developmental psychology, that the ability to reason counterfactually has a significant maturational component and does not arrive at adult levels of competence until about 12 years of age (Rafetseder, Schwitalla & Perner, 2013). However, in contrast, other scientists have discovered the ability to reason counterfactually in children as young as 4 years old (Beck, Robinson, Carroll & Apperly, 2006). It seems reasonable to conclude from the literature, therefore, that a continuum of ability to reason counterfactually exists in children between the ages of 4 and 12 years.

From the educator's point of view, what matters is when the ability to reason counterfactually can be considered *sufficiently* mature to incorporate into classroom praxis as a pedagogical instrument. Given the crucial nature of counterfactual reasoning to scientific inquiry (Kuhn, 1993), this question is of even greater interest to the science educator. In this paper, we make an attempt to introduce counterfactual reasoning into a science classroom at the middle school level and empirically evaluate the consequences.

METHODS

Sample

30 students (14 F, Average age = 10.5 yrs) from an elite English medium private school in Ahmedabad, Gujarat participated in the study. Two groups (A & B) of 15 students (7 from Grade 5, 8 from Grade 6 in Group A, and 8 from Grade 5, 7 from Grade 6 in Group B) were formed. Students were randomly assigned to the groups based on the order of appearance of their names in the attendance register (1st name assigned to A, 2nd to B etc.).

We used the scores of a recent class test in science to determine whether the two samples had varying levels of ability in the subject and/or intelligence differences. Even though the mean score for Group B (16.5/25) was slightly higher than for Group A (15.9/25), a two-sample T-test rejected the hypothesis that the two groups differed in ability (t_{28} = -0.353, p = 0.27). Thus, at least in a statistical sense, the two samples were ability-matched.

Protocol

The study was conducted on two different days for the two different grades (5 & 6), owing to their separate class schedules. Students from the same grade were divided into two groups (A & B) respectively. The two groups underwent a series of tasks (phases) during the study.

As illustrated in Figure 1, the study had 4 phases in all. Phases I, II & IV were common to both the groups, and are described in detail below. Students in group B underwent an additional Phase III which consisted of a counterfactual reasoning task.

In Phase I, students were shown four physical models (M1, M2, M3 & M4) of plants and animals -2 plants (cactus, water hyacinth), 2 animals (earthworm, rabbit)) to individual students. They were asked to list down the number of structural features (adaptation) that they could observe. In front of each structural feature, they were also asked to write down their understanding of the functional significance of that particular structure. In Phase II, students were provided with information cards on four different ecosystems (marine, alpine, desert and underground- E1, E2, E3 & E4 respectively). Students were asked to perform a match between

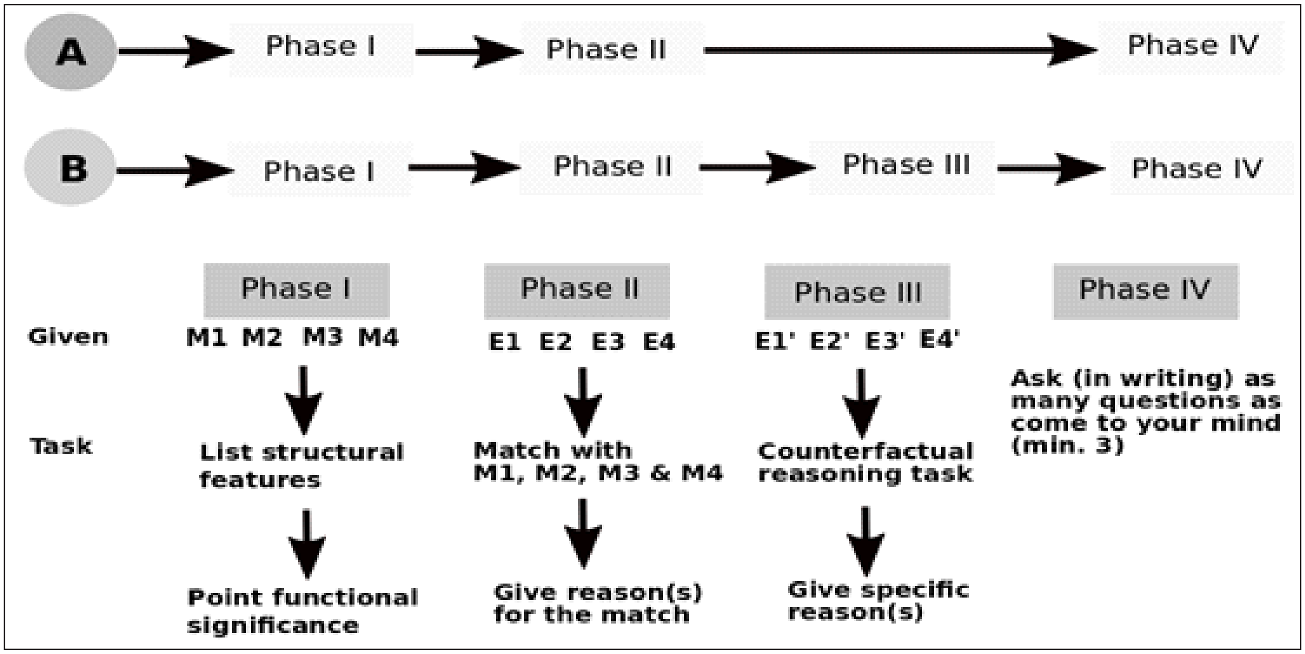

Figure 1: Study Design. See text for details

the model organisms from Phase I and the information cards provided in this phase (II) with minimum 4 reasons for their match.

Phase III consisted of the counterfactual reasoning task, which was undertaken only by Group B students. In this task, students were provided with pictures (but not models) of four new biological organisms alongside a description of the ecosystem they belong to (E1', E2', E3' & E4'), with the set of ecosystems used the same as in Phase II. Students were asked to think about specific structural modifications that would go for each organism that would make it suitable to survive in a different ecosystem. For example, if an organism "a" belongs to ecosystem A, organism 'b' belongs to ecosystem B and organism 'c' belongs to ecosystem C, then students may be asked to think about making structural changes in 'a' to make it survive in B and so on. Students were further asked to give reasons to support their structural modifications.

Phase IV, undertaken by Group A participants right after Phase II and by Group B participants after Phase III, involved a retrospective recapitulation of as many questions as the student could remember (minimum three) occurring to them throughout the activity.

Data

In this work, we present our analysis of questions posed by students in Phase IV of the study. A total of 162 questions were contributed by our 30 student sample across both groups. Interestingly, both the groups posed around same number of questions – 82 questions posed by Group A students and 81 questions by Group B. The entire set of questions were digitized, typographically (but not grammatically) corrected. Sample questions are shown in Table 1.

Group A	Group B
Why does in cactus white things come out?	Why aquatic plant leaves are broad and wax coated?
How did you get this rabbit?	Why does an earthworm have a slippery body?
Why does rabbit have a bushy tail?	What type of eyes do fish have that allow them to see in water?
Why fern look like a Christmas tree?	How do spines grow in cactus?
Why doesn't worms have legs and arm?	How do plants in the last zone of the sea survive with less sunlight?

Table 1: Examples of questions asked by both student groups

Analysis

We analyse our student-generated questions to see if priming with counterfactual reasoning task has had an impact on their quality. We categorize students' questions into multiple groups. This categorization is based upon accepted question-classification protocols reported in education literature (for instance, see Chin & Chia, 2004). We use multiple categorization protocols in order to make meaningful inference about the quality of questions asked to the effect of finding reasonable difference(s) between groups A & B, if any. Below, we report the protocols and the codes we use for categorization:

a) Chin and Chia classification: This categorization is based upon Chin & Chia (2004). We refer to this categorization protocol as the ***CC*** protocol. The categories, codes and corresponding examples from our study are given below:

(1) *Information-gathering question* ***(G)*** which pertain to mainly seeking basic factual information & whose answers are relatively straightforward, viz., 'does water hyacinth only grows in water?' (sic).

(2) *Bridging question* (***B***) that attempt to find connections between two or more concepts. For example, 'why ferns has their leaves in different rows why they can't be like normal leaves?'(sic). Here, the student is trying to link fern leaves with the concept of orientation of leaves of other plants.

(3) *Extension question* ***(E)*** which lead students to explore beyond the scope of the problem resulting in creative invention or application of prior knowledge. For example, 'why cactus store water?' (sic). Here, the student extends her prior knowledge about cactus storing water to know the reason behind it.

(4) *Reflective question* ***(R)*** that are evaluative and critical, and sometimes contribute to decision-making or change of mindsets. We use it to categorize questions which refer to some form of abstraction about a feature or function of the organism. For example, 'how water hyacinth grow in water?' (sic).

Questions that cannot be reasonably coded in any of these categories are coded ***not applicable (N)***. For example, 'why the lab(e)o fish has rhombus design on it?' (sic)

b) Chin and Kayalvizhi classification: This categorization is based upon Chin and Kayalvizhi (2002). We refer to this protocol as the ***CK*** protocol. The categories, codes and corresponding examples from our study are given below:

(1) *Investigable question* **(I),** where questions could potentially be answered by the student by following the scientific method. For example, 'what is inside fern?' (sic).

(2) *Non-investigable question* **(N)**, where questions could either not be answered, or were simply probes for factual information. For example, 'what is the white liquid inside cactus?' (sic).

c**) 5W1H Model:** Finally, a very general semantic categorization of questions - the 5W1H model (for 5 W's: who, what, when, where, why & 1 H for how), traceable all the way to classical antiquity in its provenance - can be applied in any information-gathering setting, including ours. We categorize questions as *why (Y), what (T), where (R), when (N), and how (H)* following the classical protocol, but add an extra category for questions requesting *statements of properties (S)*, e.g. 'is cactus poisonous'?

We also found that at times there were two parts to a question; one was a leading question and the other had either the 5Ws or H posed. In such cases, we coded only for the leading question. To build intuition for the relative strengths and weaknesses of these categorization protocols, we display some sample categorizations in Table 2.

No.	Question	CC	CK	5W1H
1	Why can humans live in almost all places but most animals can live only in certain habitats?	B	I	Y
2	What is the use of long horns in Arabian Oryx ?	E	I	T
3	Why ferns have so short leaves?	R	N	Y
4	Why does frog have blue blood?	E	I	Y
5	Why does the cactus have thorns?	R	I	Y

Table 2: Sample questions from both groups alongside their coded categorizations

Question 1 in Table 2 is coded a *why (Y)* question in the 5W1H protocol because of its semantic intent. It is also coded as an *investigable (I)* question in the CK protocol because the underlying premise can be scientifically investigated, unlike the premise of Question 3, for example, where the premise is a subjective value judgment (N). Question 1 is also coded as a *bridging (B)* question in the CC protocol, since the student appears to be bridging the concepts of survivability and adaptability with the question, unlike say in Question 2, where the student is extending her prior knowledge about the oryx's long horns to understand its purpose (hence coded *extension (E)* question). The code for each protocol for each question was independently coded by two researchers with background in cognitive science & biology education respectively. With an initial 80% match in the coding, the discrepancies were sorted via discussion and a consensus was reached to prepare the final code.

Statistical hypothesis testing of group effects on categorization were conducted using two-sample chi-square tests to quantify the likelihood of whether the categorizations of questions resulting from both groups could stochastically have been sampled from the same underlying discrete distribution.

RESULTS

Our primary hypothesis was that we would find differences in the nature of questions emanating from groups

A and B because of the additional counterfactual reasoning task performed by students in group B but not group A. Below we illustrate the group-level categorizations obtained via each of the three different protocols.

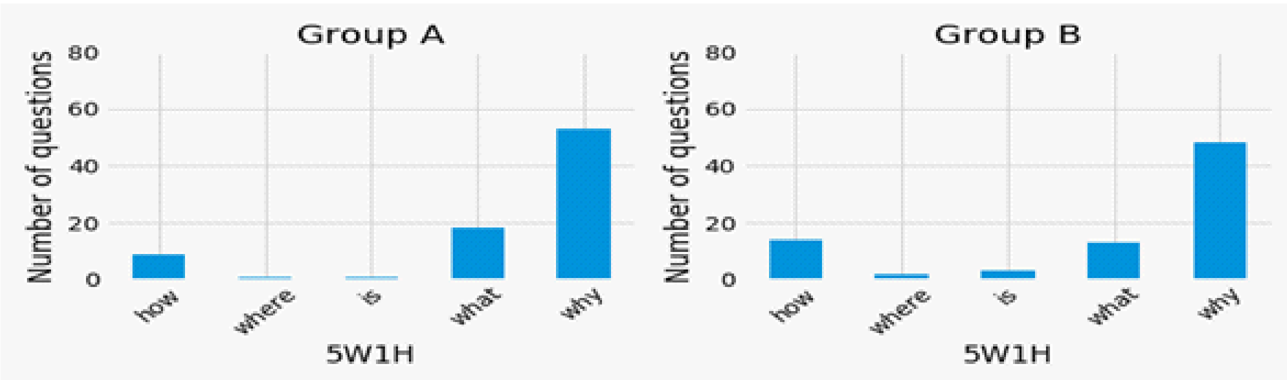

Figure 2: Categorization of students' questions using the 5W1H framework

It is both visually apparent (Figure 2) and supported by chi-square testing (χ^2 = 3.45, p = 0.49) that the semantic categories of questions asked by both student groups are virtually identical. This is a reassuring observation, since it supports the case that any differences found between the two groups will not be a function of language proficiency.

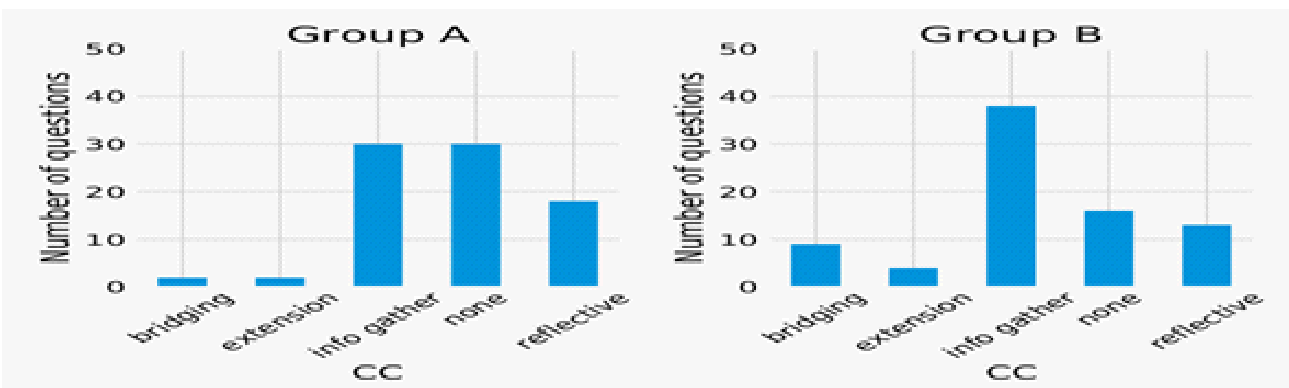

Figure 3: Categorization of students' questions using Chin & Chia's framework

Both visual inspection (Figure 3) and statistical testing (χ^2 = 11.11, p = 0.025) identify significant differences in the question categories seen using Chin & Chia's categorization protocol. In particular, the use of counterfactual reasoning appears to have stimulated the generation of considerably more questions seeking to bridge between concepts (11% of all questions for Group B versus 2.5% of all questions for Group A). Growth is also seen in information-gathering questions, with a corresponding reduction in questions that could not be placed in any category in the CC framework (37% of all question for Group A versus 20% of all questions for Group B).

However, we did not find any significant difference statistically (x^2 = 0.88, p = 0.34) in the CK protocol (Figure 4), perhaps due to the limited number of categories present in this protocol.

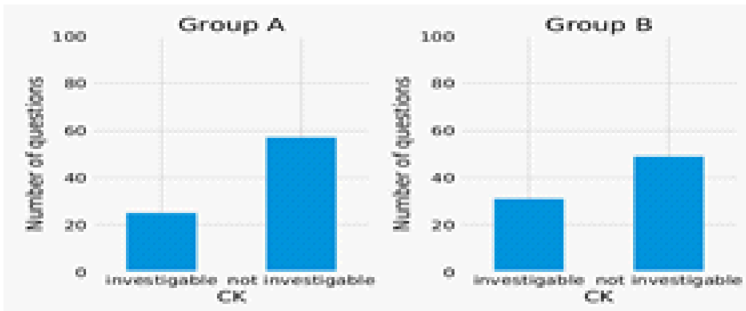

Figure 4: CK protocol Q categorization

The overall gist of our results is that there is a discernible change in the nature of questions being asked by Group B students. The counterfactual reasoning intervention used appears to have stimulated at least some of these students to ask questions seeking to '*bridge*' (see CC protocol) concepts presented in the activity material with other concepts.

Table 3 presents all eight (one question was repeated by two students) unique bridging questions asked by Group B students to bring out the fact that these questions are, in fact, interesting and likely to stimulate deeper understanding of associated concepts. Counterfactual reasoning (italicized) is evident in several of these questions, suggesting that the reasoning task given to the students directly contributed to the change in question pattern.

How can the fish keep its eyes open underwater?	Does the sea lion's nostrils close automatically when they dive into the water?
Why instead of germinating on the dark forest floor, their seeds germinate high up in the mature tree whereas tiny seedlings can get light?	*Why do sea lions propel why do not they have fins just as the fish?*
Generally fishes and aquatic animals have thin skin and less weight, but seal is heavy still is a good swimmer. How?	*Why can humans live in almost all places but most animals can live only in certain habitats?*
Does the pitcher plant have slippery surface so that insect slip inside ?	*Why do Arabian Oryx don't live in Arctic or Boreal region just as rabbit?*

Table 3: Bridging questions asked by Group B students

DISCUSSION

In this paper, we report results from an experiment seeking to identify whether the use of counterfactual reasoning as a learning device for 10-11 year old Indian school children was likely to result in measurable pedagogical benefits. Using a controlled across-subjects design, and a suite of question categorization protocols, we demonstrated a significant effect caused by the use of a counterfactual reasoning activity on the quality of questions asked by students. We also established that the change in quality is not because of semantic changes, but because students become more likely to ask questions that seek to bridge their understanding across multiple concepts.

The empirical study reported in this paper was limited both in scale and scope. Replications of our results for large samples, over longer time-scales, and using more intensive intervention strategies (multiple sessions instead of single ones, bridging multiple concept sets instead of just one set) are clear directions for future work. However, given the paucity of studies on the efficacy of counterfactual reasoning on school students' understanding of scientific concepts, the present work may serve as a stimulant for further activity.

Our results suggest that counterfactual reasoning ability is sufficiently developed in middle school students to integrate activities built around it in science curricula, potentially within the existing ambit of inquiry-based learning (Marx et al., 2004). In doing so, they are in concord with a large array of results from cognitive psychology in the current decade pointing to the sophisticated reasoning capabilities of even very young children that strongly support a strong reconsideration of how curricula and pedagogy for pre- and middle school should be conducted (Gopnik, 2012).

While it is beyond the scope of the present paper to comment on the likelihood of success of this larger ambition, we propose that our more modest ambition of including more inquiry-based learning using counterfactual reasoning in middle school curricula could be fairly evaluated and concretely implemented based on our results and downstream replications planned as future work.

REFERENCES

Beck, S. R., Robinson, E. J., Carroll, D. J., & Apperly, I. A. (2006). Children's thinking about counterfactuals and future hypotheticals as possibilities. *Child Development*, *77*(2), 413-426.

Chin, C., & Chia, L. G. (2004). Problem based learning: Using students' questions to drive knowledge construction. *Science education*, *88*(5), 707-727.

Chin, C., & Kayalvizhi, G. (2002). Posing problems for open investigations: What questions do pupils ask?. *Research in Science & Technological Education*, *20*(2), 269-287.

Gopnik, A. (2012). Scientific thinking in young children: Theoretical advances, empirical research, and policy implications. *Science*, *337*(6102), 1623-1627.

Kuhn, D. (1993). Science as argument: Implications for teaching and learning scientific thinking. *Science education*, *77*(3), 319-337.

Marx, R. W., Blumenfeld, P. C., Krajcik, J. S., Fishman, B., Soloway, E., Geier, R., & Tal, R. T. (2004). Inquiry based science in the middle grades: Assessment of learning in urban systemic reform. *Journal of research in Science Teaching*, *41*(10), 1063-1080.