# Work in Progress: Identifying and Sharpening Student's Conceptions of Induction in a Coursework Context

**Nirvaan Khera, University of Illinois at Urbana - Champaign**

Nirvaan Khera is a Graduate student in the Department of Physics at University of Illinois at Urbana - Champaign. He is a Physics Education Researcher, interested in student conceptual resources and coherence between classroom contexts and everyday thinking.

**Chandra Chekuri, University of Illinois at Urbana - Champaign**
**Prof. Lance Cooper, University of Illinois at Urbana - Champaign**

S. Lance Cooper is Professor and Associate Head for Graduate Programs in the Department of Physics at the University of Illinois at Urbana-Champaign. He received his B.S. in Physics from the University of Virginia in 1982, his Ph.D. in Physics from the University of Illinois in 1988, and he was a postdoctoral research associate at AT&T Bell Laboratories from 1988-1990. His research interests include optical spectroscopic studies of novel magnetic and superconducting materials at high pressures, high magnetic fields, and low temperatures. Since 2013, he has co-taught (with Celia Elliott) a graduate-level technical writing course each spring to physics and engineering graduate students.

**Eric Kuo, University of Illinois at Urbana - Champaign**
**Yael Gertner, University of Illinois at Urbana - Champaign**

Dr Gertner joined the Computer Science Department at the University of Illinois in 2020 as a Teaching Assistant Professor. She received her B.S. and MEng in Electrical Engineering and Computer Science from MIT, and Ph.D. in Computer and Information Science at the University of Pennsylvania. She was a Beckman Fellow at the University of Illinois Urbana-Champaign. Her current focus is on broadening participation in Computer Science and Computer Science Education She has been developing materials and teaching for iCAN, a new program for broadening participation in CS for students who have a bachelor's degree in a field other than computer science.

# Work in Progress: Identifying and Sharpening Students' Conceptions of Induction in a Coursework Context

## Introduction

Mathematical induction is an important and difficult topic in the Computer Science curriculum [1]. In order to effectively teach mathematical induction in large courses we need to understand how student ideas evolve when interacting with course content and worksheet style coursework. However, prior research has largely focused on gaps in student understanding rather than how coursework can fix it[2,3,4,5]. Notable exceptions include works of Harel [6] and Norton et al [7]. We extend their effort by analysing 9 think-aloud interviews to better understand changes in an individual's thinking in response to worksheet-style questions followed by socratic-style questions from the interviewer. The socratic-style questions also serve to be a model for future worksheet style questions with the goal of having students elaborate on their ideas[8]. The interview aims to both elicit student ideas, as well as be an emulation of worksheet style questions. To this end, we ask the following research questions:

1. What are the ideas students bring in solving mathematical induction tasks, and how do they add to prior literature about gaps in student understanding?
2. How do these ideas evolve as the interview progresses? Can students refine their knowledge as a result of the socratic-style questions?
3. What aspects of the interview enable these ideas to evolve?
4. Does such an interview allow itself to be emulated by worksheet-style questions?

In this work in progress we focus on Research Question 1. Questions 2,3,4 will be topics of future research. To answer research question 1, we take a resources framework of student ideas [9]. We believe that even in canonically incorrect answers, students bring pieces of knowledge that have the potential to be useful in forming correct answers. For example we find that even in the case of a student believing base cases are not essential to an inductive proof, they still reason that a base case has the auxiliary function of determining if the claim is true to begin with before attempting induction. While this idea does not acknowledge the role of a base case in a formal inductive proof, it represents a practical understanding of a way finding base cases help in producing an inductive proof.

Prior work has found a range of incorrect ideas students bring to mathematical induction that instruction needs to target, such as: believing base cases are not essential to the proof [2], believing induction to be circular reasoning [3], being unable to appreciate the role of the principle of universal generalization in an inductive proof. [7]. We try to explore whether these findings are replicated and if so, do these ideas contain productive resources that can then be leveraged to improve student understanding.

Interview protocol

Question 1**:** we focus on whether students can do induction in the familiar context of proving a summation identity using Mathematical Induction, and explain why it works. The students were asked the following question.

1) Evaluate the following sum:

$$\sum_{i=1}^{25} \frac{1}{i(i+1)}$$

Even though the induction hypothesis is not given, we give hints to make it easy to find the pattern. Hints include revealing the final answer, and/or directing students to calculate the first 3 partial sums. The sum from 1 to n gives the easily recognizable form of $\frac{n}{n+1}$. The goal here is to see whether students can do mathematical induction and answer why induction works.

Question 2: we present the following problem and then proceed to give students a correct proof.

2) Juice is sold in either a small box containing 3 packs or a large box containing 5 packs. You can only buy complete boxes, but you can buy any number of boxes of each type. For example, if you wanted 11 juice packs, you could buy a large box containing 5 and two smaller boxes containing 3 packs each.

Consider the following proof written by a student. The student shows that for any n ≥ 8, one can get n packs exactly by buying boxes with 3 or 5 packs each.

This is followed by a correct textbook-style proof given to the students. The proof includes showing the base cases 8,9,10 and a strong induction step showing that assuming P(k) is true for all $8 \leq k < n - 1$ implies that P(n) must be true.

The correct proof is followed by asking whether the student thinks the proof is correct. Then, we ask questions about the choices made in the proof such as the choice of the 3 base cases, using n ≥ 8 in our claim, and the wording of the inductive assumption. We also then ask whether students can prove a similar claim if the box sizes were 4 and 5 juice packs respectively. We take the ability to solve this question as a signal of understanding the correct proof given to them.

The goal of this question is to ask questions about induction in the less familiar context of having multiple base cases which also do not start with n=0 or n=1,. Additionally, we want to document the productive resources students bring to understanding a textbook style proof.

## Methods

We interviewed 9 students. 3 were from the introductory algorithms honors course (Francis, Gray, Iris), 4 were from the introductory data structures honors course (Alex, Brian, Drew, Harry) , and 1 was from the regular algorithms course (Ellise) and one was doing an upper division proof based course and had already taken the algorithms and data structures courses (Carlie). We focused primarily on honors students to see whether difficulties in articulation also exist within the students with the highest efficacy. We recruited students by making in class announcements in the Discrete Math honors and the Data Structures honors courses at our university. We interviewed all 9 students that agreed to schedule an interview. Each interview lasted 90 minutes and was audio and video recorded with the interviewer, interviewee and their worksheets being visible.

To analyse the data, the interviewer created transcripts of each interview, and created a table of quotes that best illustrated how students answered questions 1 and parts of question 2 that saw the most variety in student responses(why do we need 3 base cases, and what if the juice boxes contained 4 packs and 5 packs). Then we summarized the range of responses to these questions, noted how many students shared the dominant ideas, and selected some quotes that best illustrated the group of ideas. In future work, we will further analyse how student answers changed through the interview

## Findings

We find that even though a majority of students are able to write a proof of the identity in Q1, some students struggled with answering conceptual questions about why induction works, and answering questions about the non-conventional Q2. Q1 is not an especially easy standard induction question, and yet ability to do it does not necessarily mean being able to explain how induction works. This suggests that even though students can procedurally solve standard induction questions, they struggle with canonical explanations of why induction works. However student explanations still show promising resources that can be leveraged to develop better conceptual understandings. We document the difficulties students have for answering each question and highlight the productive resources they bring to them. We believe knowing the resources and the difficulties together will help design material to leverage the resources to address the difficulties.

Question 1:

We see that out of 9 students, all were able to write a proof by induction for Question 1. 6 out of 9 were able to give an explanation of why induction works, and 5 of those are in line with canonical explanations.

For example, Gray and Francis explain that induction works by creating a chain of implications from an arbitrary n. Making it so the truth of P(n) depends on the truth of P(n-1), which depends on P(n-2) and so on. This chain of implications goes on until the base case for which we know the proposition is true and so we can conclude its true for the entire chain of implications and therefore P(n). Alex and Brian had a similar explanation but instead of going backward from P(n), he says the base case ensures P(1) is true and the inductive step then ensures that P(2) is true, which ensures P(3) is true and it ‘propagates’ forward. Carlie focuses her explanation on the recursive nature of the inductive step. She says if you have already shown that the expression is true till x, and you can use the inductive step to show x+1, then you can recursively show x+2, x+3 and so on.

Iris gives a non-standard explanation. She argues that there is a pattern in the process of adding the 1/i(i+1) to get from n equals 1 to 2 and from 2 to 3. She argues the pattern would hold when going from any k to k+1. She argues that the inductive assumption was created from this pattern and therefore believes it must hold true. In Literature, this explanation is called Process Pattern Generalization [6].

3 students, Harry, Ellise and Drew had some difficulty explaining the logic of induction. Drew linked the truth of any P(k) to P(k-1) and the truth of P(k-1) to P(k-2) and so on but didn’t present any argument of it ending up in the base case. He also showed sophistication to say that induction only applies on certain sets of numbers like the natural numbers but also thought it worked on real numbers. When asked why a base case was essential to an inductive proof, Ellise said “I want to say yes but I cannot think of why”. This is despite the fact that Ellise grasps some of how the induction step works. She says that for example “if my k is 5 and I prove that it is true for 6, then I could have assumed my k was 4 and I have shown that k is true for 5”. Harry said the base case ‘roots’ the proof in an ‘actual value’, but he was unable to produce further explanations when asked to elaborate. This explanation has the appropriate elements but needs to be elaborated on.

Question 2:

Question 2 has non conventional features for an induction problem, such as having 3 base cases not including n=0 or n=1. You also need to use the truth of n-1,n-2,n-3 to show n instead of just n-1 as it is with conventional summation questions. We see that this leads to some troubles to even the students who were able to explain the process of induction for Q1 very well.

While all students agreed that the proof was correct, only Brian, Carlie, Francis, Gray and Iris gave correct explanations to the reason behind the choices made in the proof. We will focus on answers to why we choose 3 base cases since it discerns their ability to answer most of the other questions.

Here is an example of a canonically correct answer:

> *Carlie: The reason behind 3 base cases is that 3 is the smallest interval you can move forward by. You can technically prove it for 5 but that is kind of overkill. But either works. 3 is easiest to find 3 in a row. If you know the number three numbers ago followed a property you can just add 3 and end up at the number you are at right now. So if basically at any given point in time you know that 3 numbers ago the property held you add just 1 more small pack and you get your number.*

She describes that the important detail is that 3 is the smallest of the 2 box sizes, and if you can construct a number from large and small boxes, you are also guaranteed to be able to do so for a number 3 greater by just adding 3 (adding a small box). Brian, Francis, Gray and Iris had similar answers.

Four students were unable to answer this. Drew mentions that the number of base cases corresponds to the 3 chosen when defining n'=n-3, and posits incorrectly that you can choose n'=n-1 and it would correspond to requiring 1 base case.

Alex gives the following explanation.

> *Alex: now I need to show that there will be a situation in which I am only buying only small, only large or only small plus large, which is exactly why they have 3 base cases here, n=8 corresponds to some combination of small and large, n=9 is only small, and n=10 is only large. This is why we need 3 base cases.*

This logic is incorrect and it leads Alex to later claim that if the question asked for a proof of $n \geq 12$, then he would need the base cases of 12,13,14,15 because 13 and 14 are both a combination of small and large so we need another base case to cover the case with only large boxes. Harry offers the same explanation though he also comes up with the correct reasoning later in the interview. Ellise offers a correct explanation of how we utilize the truth of 8,9 and 10 to prove the claim for the following numbers but still believes that the base case is auxiliary to the rest of the proof.

Students were then asked a variant of the juice box problem that pertained to creating and proving a claim for which n can be constructed with juice boxes of 4 and 5 respectively. Drew and Alex didn't reach this question in the protocol. Everybody recognized the need for 4 base cases. Brian, Gray, Iris did not construct a claim with the right bounds. They said that the claim is true for $n \geq 9$ possibly extrapolating the pattern from the 3 and 5 juice box question. For example Brian says we would need 9,10,11,12 as base cases but does not check whether each is true. Carlie, Harry and Ellise started by assuming the bound will be $n \geq 9$ but then caught that 11 doesn't work and corrected the bound. Iris found that $n \geq 11$ doesn't work and said that it might not be true that there is any bound for which the claim holds. Francis found the correct bounds.

Conclusions

We find that students in our sample are largely well equipped to do Mathematical induction in the familiar context of proving a summation identity, but encounter some difficulties when dealing with a non-conventional situation such as the juice box problem which doesn't involve mathematical expressions, and has a non-trivial number of base cases. However, even when challenged, students are able to form productive ideas, and sometimes change their explanations over the course of the interview.

Our findings are heavily based on responses of honors students. These students are likely to have higher efficacy than the broader CS student population. The finding that even honors students struggle with Q2 despite having excellent answers to Q1 suggests that the inability to reason about Q2 is likely also present in the broader CS student population.

Future Directions:

In the future we are interested in further examining and categorizing the productive ideas that students bring while solving induction problems.

We are also interested in how these productive ideas form and change over the course of the interview. By studying these changes, we want to understand what aspects of the interview encouraged these ideas to change, and whether it can be incorporated into worksheet style questions. For example which questions/subquestions or interviewer moves push students to elaborate on or change their ideas

In the future we want to interview students who are not in honors classes. We are also interested in analysing more questions from our interview protocol not included in this paper.

References


[1] K. Goldman *et al.*, “Identifying important and difficult concepts in introductory computing courses using a delphi process,” *Proceedings of the 39th SIGCSE technical symposium on Computer science education*, Mar. 2008, doi: https://doi.org/10.1145/1352135.1352226.

[2] G. J. Stylianides, A. J. Stylianides, and G. N. Philippou, “Preservice teachers’ knowledge of proof by mathematical induction,” *Journal of Mathematics Teacher Education*, vol. 10, no. 3, pp. 145–166, 2007.

[3] N. Movshovitz-Hadar, “The false coin problem, mathematical induction and knowledge fragility,” *Journal of Mathematical Behavior*, vol. 12, no. 3, pp. 253–268, 1993.

[4] J. Segal, “Learners’ difficulties with induction proofs,” *International Journal of Mathematical Education in Science and Technology*, vol. 29, no. 2, pp. 159–177, Mar. 1998, doi: https://doi.org/10.1080/0020739980290201.

[5] I. Polycarpou, “Computer science students’ difficulties with proofs by induction,” pp. 601–606, Mar. 2006, doi: https://doi.org/10.1145/1185448.1185579.

[6] G. Harel, “The Development of Mathematical Induction as a Proof Scheme: A Model for DNR-Based Instruction”.

[7] A. Norton, R. Arnold, V. Kokushkin, and M. Tiraphatna, “Addressing the cognitive gap in mathematical induction,” *International Journal of Research in Undergraduate Mathematics Education*, vol. 9, no. 2, pp. 295–321, 2023.

[8] A. D. Robertson, L. M. Goodhew, L. C. Bauman, A. T. Alesandrini, and P. R. L. Heron, “Developing and assessing research-based, resources-oriented instructional materials: Attending to conceptual resources in (ACORN) physics tutorials,” *Physical Review Physics Education Research*, vol. 21, no. 2, Oct. 2025, doi: https://doi.org/10.1103/zsqv-4py5.

[9] D. Hammer, A. Elby, R. E. Scherr, and E. F. Redish, “Resources, Framing, and Transfer,” *Transfer of Learning from a Modern Multidisciplinary Perspective*, pp. 89–120, Sep. 2000, doi: https://doi.org/10.1108/978-1-60752-673-520251004.