Int. J. Res. Undergrad. Math. Ed.
https://doi.org/10.1007/s40753-022-00163-2

Addressing the Cognitive Gap in Mathematical Induction

Anderson Norton[1] · **Rachel Arnold**[1] · **Vladislav Kokushkin**[1] · **Marcie Tiraphatna**[1]

Accepted: 4 January 2022

Abstract

Proof by mathematical induction poses a persistent challenge for students enrolled in proofs-based mathematics courses. Prior research indicates a number of related factors that contribute to the challenge, and suggests fruitful instructional approaches to support students in meeting that challenge. In particular, researchers have suggested quasi-induction as an intuitive approach to understanding the role of the inductive implication. However, a cognitive gap remains in transitioning to formal proof by induction. The gap includes the cognitive demand of quantification and the danger of inadvertently assuming what one is trying to prove. Informed by prior research, we designed instruction that builds from students' conceptualizations of logical implication, utilizes quasi-induction, and introduces a novel set of tasks designed to bridge the gap that remains. We studied this research-based instructional design within two sections of an Introduction to Proofs course at a large public four-year university in the southeastern United States. Our findings, presented as themes, bring together extant literature and highlight nuances not previously reported, particularly regarding reasons students appear to conflate the assumption of the statement to be proved with the assumption of the inductive hypothesis. For instance, quantification of the inductive implication and the inductive assumption require careful use of language, using terms that may seem ambiguous to students, outside of mathematical convention. We conclude with a discussion of links between the Principle of Universal Generalization and mathematical induction.

Keywords Advanced mathematical reasoning · Research-based instruction · Mathematical induction · Proof and proving · Principle of universal generalization

✉ Anderson Norton
norton3@vt.edu

[1] Department of Mathematics, Virginia Tech, Blacksburg 24061-0123, VA, USA

Published online: 07 March 2022

Introduction

Prior research has identified several challenges experienced by students when learning proof by mathematical induction (MI). For example, students may overlook the role of the base case (Baker, 1996; Ernest, 1984; Ron & Dreyfus, 2004; Stylianides et al, 2007) or they may conflate the assumption of the inductive hypothesis with the assumption of proposition that is to be proved (Avital & Libeskind, 1978; Movshovitz-Hadar, 1993).

We review prior research on such issues and other factors contributing to the challenge of mastering proof by MI. Then, we introduce research-based instructional approaches to support students in overcoming this challenge. Specifically, researchers suggest naïve induction, or quasi-induction (QI), as an intuitive, informal approach to understanding the role of the inductive implication (Avital & Libeskind, 1978; Harel, 2001). However, Harel (2001) identified a cognitive gap as students transition from QI to formal inductive proofs. To understand this gap, we studied two classrooms that leveraged research-based instructional approaches. Results of the study address the following research questions:

1. Following research-based instruction on MI, what cognitive gap remains as students formalize the proof technique?
2. How do students experience challenges identified in prior research on proof by MI? What reasoning do students demonstrate as they address this gap?

Intellectual Need and Challenges of Proof by MI

Students' experience of mathematics is often very different from that of teachers. Dawkins and Weber (2017) assert that students sometimes experience learning difficulties because they are encouraged to adopt mathematicians' expectations in the production of mathematical knowledge without personally assigning the same importance. This discrepancy is closely tied to students' intellectual need for formal mathematics.

According to Harel (2008), intellectual need is a learner's subjective need to address a problem by learning something new. Cognitive challenges naturally arise in students' attempted production of new knowledge. In this section, we summarize prior research on the cognitive challenges that students experience in learning proof by MI.

Proof by MI poses notoriously persistent challenges for college students (Baker, 1996; Michaelson; 2008; Stylianides et al., 2007). Researchers have found that, oftentimes, learners can mechanically generate an inductive proof without developing a robust understanding of its methodology (Knuth, 2002). Movshovitz-Hadar (1993) studied 24 pre-service secondary mathematics students enrolled in a problem-solving course at an Israeli university. Findings suggested that instruction led to procedural knowledge for completing inductive proofs without an underlying logical justification. Therefore, when students faced situations of cognitive conflict, their confidence was shaken. One participant noted, "I never really believed in [MI] as a method of proving conjectures, although I know quite well how to do it, technically I mean" (p. 262).

Researchers have identified several factors that complicate proof by MI. These factors include the role of the base case (Baker, 1996; Ernest, 1984; Ron & Dreyfus, 2004; Stylianides et al., 2007), domain-specific knowledge particular to the proposition (Baker, 1996; Davis et al., 2009; Dubinsky, 1991), understanding logical implication as an invariant relationship (Dubinsky, 1986, 1991; Norton & Arnold, 2017), and the importance of (hidden) quantifiers and use of related language (Ernest, 1984; Movshovitz-Hadar, 1993).

Regarding the base case, Ernest (1984) noted students often treat its verification as a formality without understanding its essential role in MI. This treatment aligns with many students' and teachers' procedural treatment of MI as a list of steps to be checked (Stylianides et al, 2007). When interviewing teachers, Ron and Dreyfus (2004) found that teachers sometimes treated the base case as an example of the conjecture rather than the starting point for MI. Dubinsky (1986, 1990) noted one of the greatest difficulties for students in mastering proof by MI stems from their understanding of the base case as a formal step without real importance. Harel (2001) found similar issues within learners' perceptions of the base case: "when pressed, some students admit that they view the step of verifying $P(1)$ as non-essential, and one is required to perform it just to satisfy the rule stated by the teacher" (p. 14).

Regarding domain-specific knowledge, researchers have noted the mathematical content of the proposition itself may compound students' struggles with MI. This is especially true as students attempt to prove the inductive implication. Not only must students maintain their understanding of the implication's logical role in MI, they must also wrestle with the mathematical relationship it describes. Davis et al. (2009) documented pre-service teachers' struggles as they attempted to prove an inductive implication involving powers of 3. One pre-service teacher noted, "you need to know a lot of background algebra knowledge and different mathematics knowledge, like 3^{n+1} is 3^n times 3" (p. 6).

The remaining two factors – logical implication as an invariant relationship and the role of quantification – play a more prominent role in our study. We elaborate on research related to these factors below.

The Inductive Implication as an Invariant Relationship

The inductive implication $P(k) \rightarrow P(k+1)$ can be treated in two ways: (1) as an inductive step, $P(k+1)$, from the inductive assumption, $P(k)$; or, (2) as an invariant relationship between $P(k)$ and $P(k+1)$ for any k (Dubinsky, 1986). This distinction parallels one noted by Knuth et al. (2006) regarding the equal sign as an operator or as an invariant relationship. Dubinsky (1991) hypothesized that success in proof by MI hinges on understanding logical implication as an invariant relationship – "the encapsulation of the process of implication" (p. 111; see also Dubinsky, 1986).

In his conceptual analysis of proof by MI, Ernest (1984) distinguished two aspects of the inductive implication that might present challenges for students: (1) understanding implication as a binary function wherein P(k) is transformed into P(k+1); (2) proving the implication (deriving the inductive step from the inductive assumption). The first relates closely to Dubinsky's (1991) concern about the treatment of logical implication as an invariant relationship. The second depends more on the particular context of the conjecture that is to be proved. Empirical studies have held up both aspects of conceptualizing logical implication as critical to understanding MI.

One such study reported on the "knowledge fragility" of pre-service teachers (Movshovitz-Hadar, 1993). Although many students demonstrated confidence and procedural knowledge for completing proofs by MI, they struggled to conceptualize the method. Movshovitz-Hadar noted students apparently conflated the inductive assumption with the conjecture they were proving, as evidenced by comments like "in [MI] we always assume what we need to prove, then we prove it" (p. 262). Students were not treating the inductive implication as a single logical component that must be validated. Rather, they were focused on the inductive assumption as its own entity whose validity was taken for granted.

Palla et al. (2012) identified similar challenges among high school students in Greece. Surveys from 213 eleventh-grade students demonstrated they held incomplete definitions of MI, often including what the researchers refer to as "circular reasoning" in their description of the inductive implication. This reasoning often resulted from students' step-by-step procedure that separated the inductive assumption from the inductive step, without the holistic logical structure of the inductive implication. One student provided the following definition of MI:

> *1st step: I examine if it holds for the smallest possible value of n.*
> *2nd step: I claim that the proposition $P(n)$ is valid.*
> *3rd step: I substitute $n+1$ for n and if I will show that the proposition $P(n+1)$ is valid the proposition $P(n)$ holds $\forall n \in \mathbb{N}$.*
> (p. 1033)

In separating the inductive assumption into its own step, the student inadvertently assumed the conjecture itself. Follow-up interviews with 18 students indicated they struggled to derive the inductive step from the inductive assumption, echoing previous findings from other studies (e.g., Avital & Libeskind, 1978; Fischbein & Engel, 1989). Even when students explicitly referred to the domino metaphor described below, it offered little help for their completion of a proof by MI. Both findings– students' use of circular reasoning and their struggles to derive the inductive step– point to a disconnect between the inductive assumption and the inductive step that metaphors for recursion do not resolve.

Ron and Dreyfus (2004) demonstrated the limitations of metaphors for recursion through interviews with six Israeli high school teachers. Consider the domino metaphor: "Dominoes are arranged in such a way that if we push the first one it falls and causes a chain reaction– every domino tile knocks down the next one and as a consequence all the dominoes in the row fall" (Ron & Dreyfus, 2004, p. 115). Five of the six teachers used this metaphor in instruction on MI. However, while teachers recognized the recursion in the metaphor, they did not exploit its connection to the role of the inductive assumption in proving the inductive implication.

Quantification

Quantification is an essential component of MI (Ernest, 1984) as proving the universally quantified inductive implication requires the Principle of Universal Generalization (PUG; Copi, 1954). PUG justifies that one proves a statement of the form "$\forall x \in U_x, P(x)$" by considering an arbitrary $c \in U_x$ and demonstrating $P(c)$ is true. In MI, we prove the implication $P(n) \rightarrow P(n+1)$ is true $\forall n \in \mathbb{N}$ by fixing an arbitrary k, assuming $P(k)$ is true, and demonstrating $P(k+1)$ must also be true.

Movshovitz-Hadar (1993) reported on students' difficulty with the quantification of k in proving the inductive implication. On the one hand, $P(n) \rightarrow P(n+1)$ must be proved for all n– a universal quantifier, which establishes n as variable. On the other, proving the implication requires assuming $P(n)$ for some fixed but arbitrary value $n = k$. Ignoring the former quantification leads to an incomplete argument. Ignoring the latter quantification leads to circular reasoning (Palla et al., 2012).

The challenge of quantification in proof by MI is exacerbated when the quantifiers are hidden. Shipman (2016) discussed this additional challenge, noting that oftentimes students' mistreatment of quantifiers leads to the erroneous proof of a "for all" statement by example. Within the inductive implication, both $P(n)$ and $P(n+1)$ are open statements that have no truth value until n is quantified. When we write "$P(n) \rightarrow P(n+1)$" we implicitly mean "for all $n \geq 1$, $P(n) \rightarrow P(n+1)$." Durand-Guerrier (2003) discussed the difficulties of working with implication when quantification is left implicit. Ernst (1984) recommends that teachers make this (hidden) quantification explicit so that students can attend to its significance.

Lew and Mejía-Ramos (2019) studied what mathematicians and undergraduate students perceived as linguistic conventions in proof writing. They reported levels of variation in how mathematicians and students identify the issues with mathematical language. One of three themes was that students may not fully understand the nuances pertaining to the introduction of mathematical objects. These nuances include the usage of an over- or under-quantified variable and the failure to instantiate universally quantified variables. When proving the inductive implication via PUG, one introduces a new mathematical object: the variable k. In the inductive assumption, students may introduce k as "for all k", "for any k", "[fixed but] arbitrary k" or "any arbitrary k". Their choice of language may or may not indicate ambiguity in their thinking.

Dawkins and Roh (2020a, b) distinguish the role and value of a quantified variable in mastering proof comprehension. The role/value duality may be seen in epsilon-delta proofs of convergence wherein epsilon sometimes plays the role of a fixed but arbitrary value and elsewhere serves as an error bound. Likewise, in proof by MI, students make an inductive assumption; they assume that the proposition holds for a fixed but arbitrary k ($P(k)$ is true). Here, variables n and k play similar roles in referring to the argument of the proposition. However, the values of n and k differ in the way that the two variables should be quantified.

Instruction on Proof by MI

Traditionally, instructors have introduced proof by MI as a three-step procedure: (i) show $P(1)$ (base case), (ii) assume $P(n)$ (inductive assumption), (iii) deduce $P(n+1)$ from $P(n)$ (inductive step; Tsay et al., 2009). In 2008, Cusi and Malara proposed an alternative instructional approach that fosters students? understanding of proof by MI: (i) a thorough discussion of logical implication, (ii) introduction to the principle of MI through quasi-induction (QI) and the use of metaphors, and (iii) presenting examples of incorrect proofs to stress the importance of the base case. They applied this approach to a sequence of teaching experiments with 44 pre- and in-service middle school teachers in Italy, and results support claims of its effectiveness.

QI as an Instructional Approach

Avital and Libeskind (1978) suggested "naïve induction" as an instructional approach in which students demonstrate specific cases of the inductive implication. By looking across several such cases, students might reflect on the structure of the implication and realize that the relationship between $P(n)$ and $P(n+1)$ does not depend on the particular value of n, nor the truth value of $P(n)$. Harel (2001) elaborated on the instructional approach and labeled it as QI. A student employing QI shows that $P(1)$ is true and that $P(1) \to P(2)$, $P(2) \to P(3)$, $P(3) \to P(4)$, and so on. This leads to the plausible conclusion that eventually $P(n-1) \to P(n)$.

For example, consider the statement $P(n)$: *a $2^n \times 2^n$ grid of squares with exactly one square removed can be tiled with L-shaped tiles*. A student utilizing QI first demonstrates any 2×2 grid with one square removed can be tiled by exactly one L-tile (see Fig. 1).

Next, they show how a $2^2 \times 2^2$ grid with one square removed may be tiled with L-tiles by dividing the grid into four quadrants and reducing each quadrant to the 2×2 case (as illustrated in Fig. 2). Thus, $P(1) \to P(2)$. Then, the student might show $P(2) \to P(3)$ by observing a $2^3 \times 2^3$ grid can be divided into quadrants of size $2^2 \times 2^2$ and applying an argument analogous to that of $P(1) \to P(2)$. This process exploits the inductive pattern.

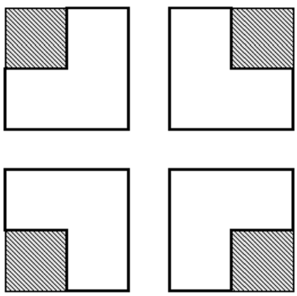


Fig. 1 The base case of the L-tiles example

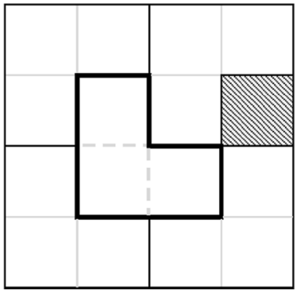


Fig. 2 $P(1) \rightarrow P(2)$

Metaphors, like the aforementioned domino metaphor, highlight the recursion inherent in MI, but research indicates that students as young as five years old demonstrate the ability to reason recursively (Smith, 2002). QI not only leverages this recursive reasoning, it draws students' attention to the role of the inductive assumption in proving the inductive implication.

The Gap that Remains

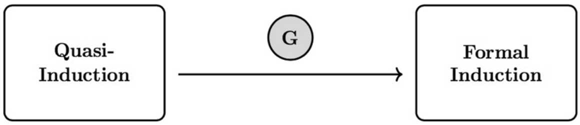
While prior research has validated QI as accessible to students and beneficial for their understanding of MI (Harel, 2001; Cusi & Malara, 2008), the conceptual jump from QI to formal MI is larger than it may seem. Harel asserts that this cognitive gap is indeed substantial, referring to formal proof by MI as an abstraction of QI. In particular, he characterizes this gap as a difference in how students perceive the inference $P(n-1) \rightarrow P(n)$. Namely, in QI, learners view it as a last step in a chain of inferences $P(1) \rightarrow P(2)$, $P(2) \rightarrow P(3)$, $\ldots$, $P(n-1) \rightarrow P(n)$ that connects $P(1)$ with $P(n)$. In contrast, formal proof by MI requires treating $P(n-1) \rightarrow P(n)$ as an invariant relationship across the entire chain.

Theoretical Framework

Our theoretical framework brings together prior research on students' cognitive challenges in proof by MI and the efficacy of QI as an instructional approach. We expand on Harel's explanation of the gap between QI and formal MI to include the tension between teachers' instructional goals and the students' experiences of the cognitive challenges that emerge from instructional interactions. In particular, our classroom study illuminates nuances in students' struggles as they begin to bridge the gap (see Fig. 3).

Our framework draws on action-object theory, which derives from Piaget's (1970) genetic epistemology, in which mathematics is understood as a product of psychology: mathematical objects arise as coordinations of mental actions through a process called reflective abstraction. In the context of MI, Dubinsky (1986) hypothesized that students' success depended on their ability to conceptualize logical implication as a single mathematical object, rather than an action that transforms the hypothesis into the conclusion. This aligns with Harel's (2001) assertion that the gap between QI and formal proof depends on students' perception of the inductive implication

Fig. 3 The cognitive gap between QI and formal MI



as an invariant relationship. While QI may be accessible to all students because it builds on students' understanding of logical implication as an action, the gap may be wider for those students who do not hold logical implication as a single object.

We employed this action-object distinction to explain difficulties inherent in proof by MI (Arnold & Norton, 2017; Norton & Arnold 2017, 2019). To better assess students' understanding of the underlying components of MI, we designed tasks (see Table 1) to be independent of mathematical content. The tasks were first used in Arnold & Norton (2017) and Norton & Arnold (2017) in clinical interviews with students from an Introduction to Proofs course. Both studies support findings from prior literature that treatment of logical implication is a major mediator in students' understandings of proof by MI (Ernest, 1984; Dubinsky, 1986). In Norton & Arnold (2017), we found that differences in understanding of logical implication may explain students' struggles with hidden quantifiers. In Norton & Arnold (2017), we report that treatment of logical implication as a single object might allow learners to re-invent a QI proof and avoid the pitfall of conflating the proposition with the inductive assumption. We note that these tasks were further refined in a follow-up study (Norton & Arnold, 2019). The results of this second study indicated that holding logical implication as a single mathematical object may enable students to attend to other components of proof by MI.

Methods

In Spring 2018, Norton and Arnold each taught a section of Introduction to Proofs at a large land grant university in the southeastern United States. Here, we describe the instruction in each class and our methods for collecting and analyzing data from the subsequent classroom interactions.

Table 1 Sample novel task

Problem Statement: Suppose $P(n)$ is a statement about a positive integer n, and we want to prove the claim that $P(n)$ is true for all positive integers n. For each scenario, decide whether the given information is enough to prove $P(n)$ for all positive integers n.

Scenario	Given Information
A	$P(1)$ is true; for all integers $k \geq 1$, $P(k)$ is true.
B	$P(1)$ is true; there is an integer $k \geq 1$ such that $P(k)$ is true.
C	$P(1)$ is true; there is an integer $k \geq 1$ such that $P(k)$ implies $P(k+1)$.
D	$P(1)$ is true; for all integers $k \geq 1$, $P(k)$ implies $P(k+1)$.
E	For all integers $k \geq 1$, $P(k)$ implies $P(k+1)$.
F	For all integers $k \geq 1$, $P(k)$ and $P(k+1)$ are true.
G	$P(1)$ is true; for all integers $k \geq 2$, $P(k)$ implies $P(k+1)$.

Research-based Instruction on MI

The instructors' approaches were similar to that prescribed by Cusi and Malara (2008), however they relied less on metaphor to elucidate the components of formal MI. Instead, the instructors used novel tasks developed in their prior research (Arnold & Norton, 2017), such as the problem in Table 1 which refers to an unspecified proposition $P(n)$. Norton used the scenarios in Table 1 to generate class discussion; Arnold used them within an informal assessment. Their tasks encouraged the use of QI, introduced components of formal MI, and attempted to elucidate the intellectual need for this proof technique. Within this approach, each instructor promoted the inductive implication as an invariant relationship between the inductive assumption and the inductive step (rather than treating them as separate components), and they explicitly addressed issues of quantification. Ultimately, similar themes emerged in both classrooms.

Norton's Class

Norton began each class with a "starter problem" displayed on the projector screen. Beginning week one (of a 15-week semester), some starter problems involved proving implications that took the form of QI, though they were not introduced as such. For example, in week 3, the starter problem was as follows: "If the sum of the first 100 natural numbers is (100×101)/2, then the sum of the first 101 natural numbers is (101×102)/2."

Norton explicitly introduced MI during the first class of week six. The starter problem prompted students to generate ideas for proving 5 divides $8^n - 3^n$. Most of class was devoted to discussing the components of MI (see Table 1). For example, when considering Scenario C, Norton asked, "what integers do we know it works for?" Student responses included "1", "1 and 2," and "1 and some k." This exchange facilitated discussion about quantification of the implication and potential conflations between the truth of the inductive assumption and the truth of the inductive implication.

The next class, Norton introduced five QI statements, like the following: "If 3 lines in the plane (none parallel and with no three intersecting at a point) form 7 regions, then 4 such lines form 11 regions." Students sat at five tables in the classroom, and each table was assigned one statement to prove and present to the rest of the class on a whiteboard. At the end of class, three students were assigned three conjectures to prove by induction. The three students presented those proofs on the third and final day of MI instruction (the first class of week seven). Much of the classroom discussion that day focused on the meaning of various quantifiers and how they might be used in MI.

Arnold's Class

Arnold introduced logical implication in the first week of class. Then, unlike Norton, Arnold gave formal instruction on quantifiers before returning to proving implications. Following prior research (Harel, 2001; Cusi & Malara, 2008), she introduced

MI (during week 9) by first engaging students in QI without mentioning the formal proof technique. Students were prompted to work in groups on the aforementioned L-tiles task by showing specific cases: $P(1)$, $P(1) \to P(2)$, and $P(2) \to P(3)$. Then, together as a class, students outlined the approach to each case.

On day two, Arnold returned to the L-tiles task, prompting students to identify the key components necessary for generalizing their argument to an "algorithm" for proving the claim for an arbitrary n. She led students to the general inductive implication, introducing the notation $P(k) \to P(k+1)$ and asking how it should be quantified. After formally stating the principle of MI, Arnold gave the metaphor of "climbing a staircase" in justifying the label "inductive *step*." Then, students were given a proof and asked to determine the claim it was proving ("for all $k \in \mathbb{Z}^+$, if $k = k+1$ then $k+1 = k+2$"). This task emphasized the role of the base case. Finally, Arnold returned to the quantification of $P(k) \to P(k+1)$, asking students whether k should begin at 1 or 2 to connect the base case with the implication.

On day three, Arnold asked students to help develop a general template for a proof by MI. This prompted a discussion of the language necessary for introducing k in the inductive assumption. Students spent the remainder of class working in groups to formally prove claims such as "for all $n \geq 0, 3 \mid \left(2^{2n} - 1\right)$."

Data Collection

The two classes met on different days in the same room. There were five circular tables in the room, each with about five students who regularly work together during class. The room was equipped with multiple cameras and microphones, controlled from an adjacent room, to capture audio-video from the whole classroom. We recorded each class during the three-day instruction on MI. Norton's class meetings were held on February 20, 22, and 27 for 75 minutes each. Arnold's class meetings were held on March 21, 23, and 26 for 50 minutes each. Recordings captured the instructors' activity, overhead projections (e.g., notes and PowerPoint slides), and students' interactions.

Data Analysis

Using existing literature, we generated an initial list of codes for aspects of MI where students might struggle (see Table 2). We used these codes in retrospective analysis of classroom data. The research team met weekly from Summer 2018 through Fall 2019 to review video recordings and code salient segments of classroom interaction. Segments of video were deemed salient if they included student-student or student-teacher interactions in which MI was discussed. Each segment was assigned at least one code and documented in an Excel spreadsheet. We used the labeling scheme "[A1 2:34]" to designate data from teacher A, class session 1, timestamp 2:34 (minutes:seconds). For example, [A2 30:30] was taken from Arnold's second class on MI, starting at 30 minutes and 30 seconds. The corresponding line used Codes G, C, and F, and included the following description: "Teacher guiding students to general quantifying, formalizing induction; student 4 suggests for all $k \geq 1$ … possibly

Table 2 Codes

Code	Meaning & Criteria	Sources
B	Base case - students check base case or discuss its role in MI	Baker, 1996; Ernest, 1984; Ron & Dreyfus, 2004; Stylianides et al., 2007, Dubinsky 1986, 1990; Harel, 2001; Palla et al., 2012
Q	Quantifiers - analyzing quantification of MI variables	Ernest, 1984; Movshovitz-Hadar, 1993
I (I_Q)	Inductive Implication (QI) - proving or applying the inductive implication	Dubinsky, 1986, 1991; Harel, 2001
D	Domain knowledge - Attending to content-specific mathematics	Baker, 1996; Davis et al., 2009; Dubinsky, 1991
C	Conflation/Comparison - discussion of truth of at least two of inductive hypothesis, implication, or proposition.	Avital & Libeskind, 1978; Movshovitz-Hadar, 1993
G	Gap - Attempts to generalize from QI to formal MI.	Emerged [N1 38:30]
R_C	Reducing problem to computational setting - attempts to circumvent logic of MI by reframing problem algebraically	Emerged [A1 35:30]
K	"Why k not n?" - introduces or questions introduction of new variable k	Emerged [A2 26:32]
F	Effects of formal instruction - referencing template for proving universally quantified statements	Emerged [A3 10:20]
N	Intellectual Need - recognizing need for formalization to replace indefinite recursion	Emerged in second phase of analysis
L	Language - selecting and analyzing appropriate language for communicating logic	Emerged in second phase of analysis

conflating cases with implication, but bothered by it and self-corrects." When no existing code fit a segment of data, we introduced a new code and then reviewed all prior videos to include the additional code, wherever applicable.

During a second phase of analysis, we organized codes around the cognitive gap. Each researcher independently reviewed instances of the gap code (G) and its connection to other codes. This analysis was guided by the overarching question, "how is each code related to the gap code, as evidenced by the line-by-line coding within teaching sessions?" To answer this question, we identified codes present in classroom interactions containing a segment coded with G.

We shared our notes in group discussions to identify the connections between codes that we saw. These connections suggested themes and instances of those themes. A byproduct of our analysis was the emergence of the codes L and N. For example, when G and Q appeared within the same classroom interaction, discussion often centered around selecting appropriate language for communicating the intended logic. Thus, the L code emerged as an extension of the G-Q connection. Having determined five overarching themes, we reviewed pertinent video recordings to identify representative examples, which are used below to elaborate on each theme.

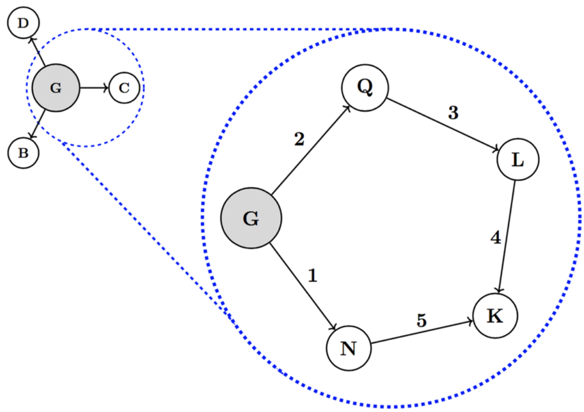
Resulting Themes

QI provides opportunities for students to engage in inductive arguments (Avital & Libeskind, 1978), but a gap remains as instructors attempt to support students' transition to formal MI (Harel, 2001). Instruction in both researchers' classrooms leveraged QI, but in different ways. Here, we describe the themes that arose from coding classroom interactions.

Prior literature has identified three main epistemological obstacles to productive teacher-student interactions regarding MI (e.g., Baker, 1996; Ernest, 1984; Movshovitz-Hadar, 1993): (1) a possible conflation between the proposition and inductive assumption (C); (2) uncertainty about the role of the base case (B); and, (3) paucity of domain-specific knowledge (D). We found evidence to support the challenging roles of the base case (B) and domain knowledge (D). We acknowledge that students rely on domain-specific knowledge to generalize QI arguments and that students must coordinate the base case with the inductive implication to complete an inductive argument. However, our analysis of the interactions centering on the gap (G) suggested the apparent conflation (C) documented in prior literature might be better understood via the connections between G and other codes.

These connections suggested the graph in Fig. 4, which illustrates an expansion of the conflation code (C) into a chain of four other codes. Nodes in the graph represent various codes identified in Table 2. Edges between nodes represent themes as connections between codes. We discuss each of these themes, edge-by-edge, in order of edge number.

Fig. 4 Graph of connected codes



Theme 1: Intellectual Need for Formal MI

Harel (2008) described intellectual need as "a behavior that manifests itself internally with learners when they encounter an intrinsic problem--a problem they understand and appreciate" (p. 285). In our study, that behavior arose from the lack of generality in QI arguments and the need for the inductive implication to continue indefinitely. Students sometimes used circular hand gestures to describe how the same implication might be used again and again.

Consider the following episode from Norton's first class on MI. The episode began as Norton transitioned from Scenario C to Scenario D (see Table 1): "$P(1)$ is true; for all integers $k \geq 1$, $P(k)$ implies $P(k+1)$." Norton is denoted N and his students are denoted X_i.

> N: So, now I know $P(1)$ is true. The only difference is here–one little change–'there exists' changed to 'for all.' Now I know $P(1)$ is true, and I know for all integers $k \geq 1$, the truth of $P(k)$ implies the truth of $k+1$. Now what do I know? What integers is the proposition true for?
> Students: All.
> N: What?!
> Students: All those positive…
> N: All I see is 1!
> X_1: But it works for all integers.
> X_2: When k is 1 and we know 1 works then $P(1+1)$ works, so 2 works.
> N: Ok, 1 works because it's working for all $k \geq 1$, so it works for 1. So working for 1 implies it works for 2. Ok, great.
> X_2: Then you can plug 2 back in for k and the logic repeats itself.
> N: This allows me to put 2 in because it says for all integers. All right, 2 is greater than or equal to 1, so this applies, and I already got $P(2)$, so $P(3)$. Great, now I'm at 3. This is gonna take a long time…
> Students: Yeaaah [laughing]
> N: How many people think this is going to work for all integers?

[All the students raise their hands.]
Student: Positive.
N: Positive integers… How do you know that there's not some integer wayyyy out there where this doesn't work?
X_3: Because this is for all.
X_1: …integers $k \geq 1$.
N: So there's a lot of conjectures y'all've been struggling to prove that kind of go on forever. Like that Fibonacci one where every third element of the Fibonacci sequence is even. And you start going and going and going, and you can't go on forever. This [points to inductive implication] goes on forever for you. This is the engine that's doing all the work for you. So that's what we want to try to do is try to apply this engine.
[N1 37:00-38:35]

Norton used the students' reasoning to highlight how the inductive implication would satisfy the intellectual need for "an engine" that would repeatedly apply the inductive argument. The students first moved from the truth of $P(1)$ to the truth of $P(2)$ using the inductive implication applied to the case where $k = 1$. Then, they argued that "the logic repeat[ed] itself" since the implication was true for all k.

A similar discussion took place in Arnold's class when students began to generalize their QI arguments for the L-tiles task. However, their arguments moved in the opposite direction, descending from an arbitrary value of n down to the base case. Arnold attempted to induce the intellectual need for generalizing QI arguments by asking students to consider the 500th case, its dependence on the 499th case, and in turn, its dependence on the 498th case. Along the way, she emphasized the importance for the base case, to which we return in a later section.

A: 498? To prove the 498th case what would I need?
Students: 497.
A: The 497th case, and to prove that case, what would I need?
Students: 496.
A: The 496th case, and it all comes down to knowing what is true?
Students: 1.
[A2 11:30]

By emphasizing the extensive effort that would be necessary to prove $P(n)$ via QI for large values of n, Arnold motivated her students to generalize their arguments to formally prove the inductive implication. Arnold is denoted *A* and her students are denoted Y_i.

Y_1: You need to know that each case for $n > 1$ can be broken down into four cases of $n - 1$ if that makes sense.
A: So, for our specific problem, we would need to show how the case that we want to show is true can be broken down into the previous case?
Y_1: Yeah.
[A2 17:47-18:26]

At times, students struggled to transform particular QI arguments into a general implication, which underscored the intellectual need for formal MI. Student X_4 in Norton's class presented a QI argument about the sum of odd numbers but did not know how to generalize it [N2 1:07:00]. He had proven that if the sum of the first six odd numbers is 6^2, then the sum of the first seven odd numbers would be 7^2. However, he did this computationally, simply adding the seventh odd number (13) to 36 (6^2) to get 49 (7^2), and it wasn't clear to him how the argument would generalize.

> X_4: This is kind of not complicated for some reason.
> *N*: Yeah it's not complicated until you start thinking how would you generalize this.
> X_4: Yeah.
> *N*: Like, did you just get lucky that 13 plus 36 gives you the next square, or is there something particular about how this 13 was constructed, 13 minus–you know, the 13 minus 1–that would make it work for the *k*th, you know, when you generalize to *k*? So I think this one's easy, but it's gonna get complicated when you generalize to *k*, and figuring out, ok, what is it about this 2 times *k* plus 1 minus 1 that's gonna lead us to the next perfect square. That's the work to do.
> X_4: Yeah unless you're flat out told that the two– that the *n*th odd number– is $2n - 1$. Then it becomes easy.
> [N2 1:07:30-1:08:23]

When students could not see the general in the particular, there was a need to generate a universal argument for the inductive implication. In the present case, this amounted to deriving $2k + 1$ as the difference of consecutive perfect squares, $(k + 1)^2 - k^2$. At other times, students provided arguments that circumvented the need for formal MI. For instance, a student in Norton's class suggested a method of descent (cf., Ernest, 1982) for proving the Fundamental Theorem of Arithmetic (FTA) that did not fit formal MI.

The FTA states that every integer greater than 1 is either prime or can be uniquely represented as the product of primes. Formal proofs of the theorem often rely on strong MI (Dawson, 2006), but the student suggested that if a given number is not prime, then it is the product of two factors greater than or equal to 2. Thus, each of those factors is no bigger than half of the original number. This repeated "halving" would eventually have to terminate so that the original number would be factored into a product of primes. When the teacher tried to redirect the class toward an inductive approach, the student protested that their approach was indeed inductive.

> N: I think that could work as a rigorous argument but that gets us into looking at limits and such. I think we can do this [gestures to board with proof idea], but I think we can come up with a shorter, cleaner proof that relies on induction instead. So that's not to discount this one [proof idea]. It's just (1) think about how induction is doing work for us, and (2), practice induction in here…
> X_2: I feel like we are doing induction with this.
> [N3: 21:28-22:45]

The student was employing recursive reasoning. The difficulty that the teacher experienced in highlighting this distinction was that the student's argument depended on a limiting process that did not terminate in an explicitly identified base case.

Theme 2: Quantifying Components of the Inductive Implication

Even when students have an intellectual need for formal MI, they may still struggle to bridge the gap between QI and formal MI. Bridging this gap entails careful quantification of the inductive implication (Ernest, 1984). Moreover, in proving the inductive implication, the inductive assumption of $P(k)$ requires a shift in the language students use to quantify k; it requires universal instantiation of k as a fixed but arbitrary integer.

The issue of quantification arose quickly in both classes despite differing instructional approaches. Arnold had provided formal instruction on existential (e.g., "there exists") and universal (e.g., "for all") quantifiers earlier in the semester. Norton, however, had not formally introduced these quantifiers prior to classroom discussion on MI. In his class, the issue arose during the first MI teaching session, as students considered the scenarios from Table 1, especially in contrasting the use of the existential quantifier in Scenario C with the use of the universal quantifier in Scenario D. Students seemed to readily understand the necessity for universally quantifying the inductive implication, but proving it was a greater challenge.

Following the discussion of the scenarios and related components of MI, students were asked to work in groups at their tables to prove the following claim using MI: "For all $n \geq 1$, $2 + 2^2 + 2^3 + \cdots + 2^n = 2^{n+1} - 2$." Each group quickly checked the base case ($n = 1$) and then focused on proving the inductive implication: for all $n \geq 1$, if the equation works for the nth case, then it also works for the $(n + 1)$st case. In proving this universally quantified implication, each group seemed to face a common difficulty. Consider the proof shown in Fig. 5, as shared by one of the groups on a whiteboard.

> X_5: We assume this "if" part [pointing to the assumption of the inductive hypothesis, at the top of Fig. 5], and then ...

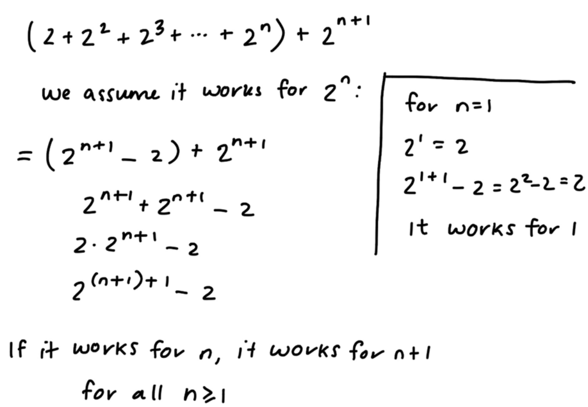


Fig. 5 Presentation of the inductive proof by first group

N: Let me ask you a question. What gave you the right to assume the "if" part?
X_5: We don't know if this is right or wrong, but we're saying if this is right, then we can get to that part [pointing to the inductive step].
N: You are not assuming the theorem. You are assuming the if part of your implication.
X_5: [laughing] Which looks a lot like the theorem.
[N1 1:03:00]

Note that this group did not explicitly quantify the inductive assumption when writing "we assume it works for 2^n." Furthermore, they used n throughout their proof, rather than introducing a new variable, k, when quantifying the implication. Figure 6 illustrates an attempted proof by a second group. This group did introduce k and attempted to quantify it.

Following the verification of the base case ($n = 1$), the proof continued with an assumption: "assume the equation holds for all $k \geq 1$ such that $2 + 2^2 + \cdots + 2^k = 2^{k+1} - 2$." It might appear that the students had assumed what they were trying to prove, but they saw the situation differently. As one representative from the group explained, "we wrote it like that, but we knew ...I guess, we didn't write the part that said, 'here is the if-then statement' because what you do immediately after that is you say 'assume that it's true." In saying this, the student seemed to think the issue was that he had not explicitly written what he was about to prove: the universally quantified implication,

$$(\forall k \geq 1)\left[P(k) \rightarrow P(k+1)\right].$$

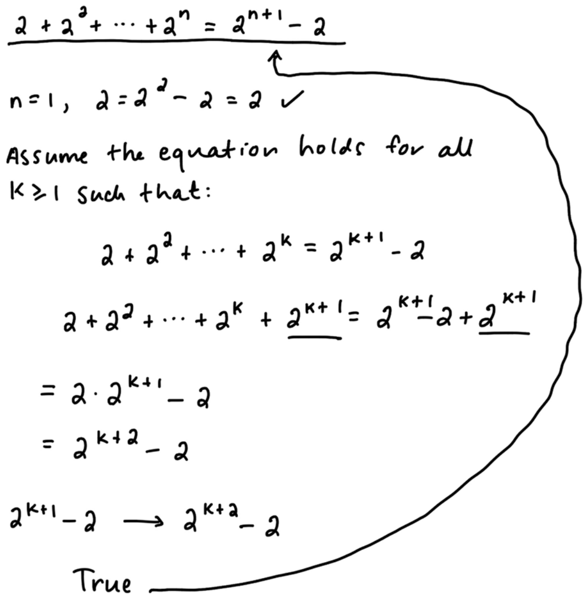


Fig. 6 Presentation of inductive proof by second group

However, Norton was more concerned that his group had universally quantified the inductive assumption,

$$\left[(\forall k \geq 1)P(k)\right] \rightarrow P(k+1).$$

On one hand, the group used "for all" language in quantifying k in $P(k)$. On the other, they also introduced "such that" language (typically associated with existentially quantified statements) before stating $P(k): \}\}2 + 2^2 + \cdots + 2^k = 2^{k+1} - 2.''$ Although Norton did not probe this use of "such that," students might have used that language to indicate that, in making their assumption, they were only concerned with values of k for which they knew the proposition to be true. The same student (X_5) used the phrase similarly in a subsequent episode.

In Arnold's class, the issue arose when she asked students what they would need to know to complete a proof by MI, building from QI arguments, such as $P(2) \rightarrow P(3)$. She had written the general implication $P(k) \rightarrow P(k+1)$ and asked how she should quantify k.

A: Someone tell me a precise way that I want to quantify.
Y_2: If $P(1)$ is true.
A: Well if $P(1)$ is true, then $P(2)$ is true. If $P(2)$ is true then $P(3)$ is true. But, how do I say that all at once so that when I go to actually write the proof, I don't have to do all those cases?
Y_3: For all $k \geq 1$, $k + 1$ is true.
A: Good [says and writes "for all $k \geq 1$"]. Say the rest.
Y_3: $k + 1$ is [brief pause] true.
A: Tell me the whole piece that I need to know, though.
Y_3: Oh, if $P(k)$ is true, then $k + 1$ is true.
A: Yes. [continues writing "$P(k) \rightarrow P(k+1)$"] If I said instead, because I know you intended to say what I wrote, but if I said this [writes "for all $k \geq 1$, $P(k+1)$ is true"] 'for all $k \geq 1$, $P(k+1)$ is true.' Let's say I was given that fact, and I also knew $P(1)$ was true. Would my claim be true, for all my positive integers?
Students: [pause, then a few students, including Y_3, respond "yes"]
A: Yeah, it would be, wouldn't it? I know that $P(1)$ is true [gesturing toward the board], and then they are saying, for every $k \geq 1$, $P(k+1)$ is true. So, that's like saying $P(2)$ would be my first value; $P(3)$, $P(4)$. This would certainly tell me that the claim is true for all things, but that is not the structure of our argument. Our argument depends on the previous case, so our argument comes from an implication.
[A2 30:10]

Arnold was asking students to quantify the inductive implication, and student Y_3 literally quantified the inductive step. However, she interpreted the students' literal response as an abbreviated attempt to quantify the inductive implication. Although we cannot know for sure whether the student made a distinction between the inductive step and the inductive implication, Arnold's interpretation seemed to resonate with him, as indicated by him nodding his head in agreement.

Later in the episode, Arnold asked the class how they would prove a universally quantified implication, and like the second group in Norton's class, they began with the language "for all."

> *A*: So, for us, we're proving a "for all." We're going to say for all k in $\mathbb{Z}$, and that if $P(k)$ is true then $P(k+1)$ is true. So let's think about how we might write that out. How would I introduce the fact that this is universally quantified? [Silence] If I translated that into a sentence, I would say ...
> Y_2: for all k...
> *A*: I wouldn't...When I prove a "for all" statement, do I say for all values all at once?
> Y_2: No, you let k in $\mathbb{Z}$ be arbitrary.
> [A2: 39:30]

The student suggested using the same language ("for all") to quantify the assumption of the the inductive hypothesis that had been used to quantify the inductive implication. Upon further questioning, he revised his response to suggest that the k used in the assumption should be an arbitrary integer. This distinction is critical because it avoids the logical fallacy of assuming what needs to be proved while maintaining a universality in k. It relies on PUG (Copi, 1954; Dawkins & Roh, 2016; Dawkins et al., 2020).

Theme 3: Interpreting the Language of Quantification

As we saw in the previous section, quantification requires careful use of language. Apparent conflations between the proposition and inductive assumption identified in prior literature, might be better described as difficulty in selecting precise, accurate language to quantify the inductive implication and the assumption of $P(k)$ in proving this implication. In our study, students had little difficulty understanding the difference between "there exists" and "for all." As one student explained, "you might think that it generalizes for all k, but we only know that it works for this one specific k that exists." [N1 36:00]. However, applying that language to various components of MI was sometimes problematic–a problem that was exacerbated by the introduction of more ambiguous terms, such as "for some." These issues came to a head during a discussion of such terms in Norton's class.

Norton invited the class to critique a quantification he had shared on a previous day: "If a proposition holds for some $k \geq 1$, then the proposition also holds for $k+1$."

> *N*: Is "some" too weak of an assumption, or is "for all" too strong of an assumption? How do I balance that?
> X_6: I think "for some" is too weak, because if you assume "for all," you're not assuming that it's true for every $k \geq 1$. You're assuming that if it was true for k, then it would be true for $k+1$.
> X_3: What if you said "any" instead?

N: What if I said "any?" I could have said, "for some $k \geq 1$"; I could have said, "for all $k \geq 1$"; or I could have said "for any $k \geq 1$." What do we think about those three? We've got "all," we've got "some"; we've got "any." Do more than one of them work? Does only one of them work? Do none of them work?
X_7: [raises hand] Um...uh...I thought I had something, but I don't. So, nevermind.
[X_5 raises hand]
N: I want to come back to X_5 in a minute, but I want to get more people into the conversation.
[long pause with silence, about 10 seconds, then one more student raises hand]
N: I'm going to make people vote.
X_3: We're going to vote?
N: Yeah, this is proof by democracy.
[N3 1:01:25]

Norton drew a chart on the board, laying out three terms: "for some," "for any," and "for all," and asked students to vote on whether each was "too weak," "too strong," or "right on." Student responses are illustrated in Fig. 7. We can discern a few general patterns from their votes. Most notably, all student responses aligned with the idea that "for some" is the weakest quantification, and "for all" is the strongest, with "for any" somewhere in between. Although students agreed about the relative strength of these terms, they widely varied in their interpretations of these terms' absolute strength and in their determination of which term should be used in the given statement. For instance, 8 out of 15 students who responded saw "for any" and "for all" as different quantifications.

In the above transcript, X_6 thought the quantification "for all" was appropriate because it was conditioned by the "if" statement. However, we reiterate that such language moves the quantification inside the implication:

$$\left[(\forall k \geq 1)P(k)\right] \rightarrow P(k+1).$$

Another student thought "for some" was sufficient because, for them, it referred to a fixed but arbitrary value. Thus, we find two issues with which students must contend in adopting language for quantification: discerning the language used to universally quantify the inductive assumption from that used to universally quantify the implication (as noted before), and interpreting terms with some agreed upon convention (cf., Lew & Mejía-Ramos, 2019).

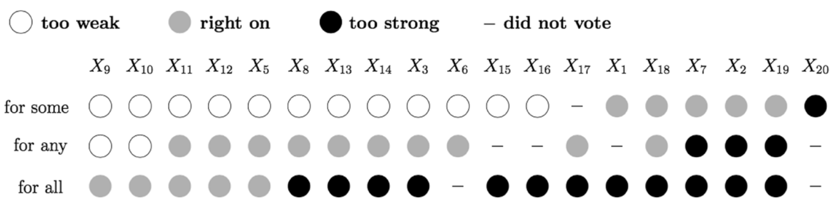


Fig. 7 Responses to Norton's class poll on quantification

Themes 4 & 5: Need for Distinction between *k* and *n* in Supporting the Language of Quantification

When introducing formal MI, many textbooks use two variables, n and k (e.g., Hammock, 2018). Although the shift from n to k may seem arbitrary to students, we have already seen how it might address ambiguities in the language of quantification. Specifically, k plays the role of a "fixed but arbitrary" value of n that leads to universal quantification by way of PUG. In addition to the intellectual need for the shift from QI to MI, we found in both classes the intellectual need for a shift in the language of quantification (Theme 4). A shift in notation, from n to k, might satisfy this need (Theme 5).

Arnold introduced the variable k during the second lesson on induction, while listing the two components of a formal inductive proof: (1) $P(1)$ is true; and, (2) $P(k) \rightarrow P(k+1)$. She motivated the need for k in the following way:

> *A*: Now, the reason I use a k here is because, if you notice, we use this implication over and over and over again, for very specific values of k. When we wrote our original outline here, we needed it to show $P(1)$ implies $P(2)$. We also needed to show that $P(2)$ implies $P(3)$. So notice that k is actually varying, because I'm going to be using this implication over and over and over and over again. Every time I want to recover the next value of n, I use this implication. I take what I already knew to be true in the previous case to recover the case that I'm looking at. So the question is, if I'm going to use this k like this, what is k? So, in other words, for which k must this be true?
> [A2 36:30]

She then asked students to discuss the issue at their tables. As the students regathered for a whole-class discussion, one said she was confused as to why they were using k instead of n. When the teacher explained it allowed them to specify fixed but arbitrary values of n, the student responded, "I just thought we were using n and k, like they were two different things." This brief episode indicates that students might not understand how k is being used to formally quantify the inductive assumption. Likewise, in Norton's class, students questioned the role of k in quantification, despite his explanation.

> *N*: So, X_3 is saying that we need two things. We need to know that it works for 1. Check, we're done with that. And she's saying that we also need to know that working for n, implies that it works for $n + 1$.
> X_8: For all n.
> *N*: Yeah, for all n, if it works for n, then it works for $n + 1$. Thank you. The "for all" is really important there, as we just saw...Now, I'm not going to get really picky about this, but some people will want you to change the letter here, because with n we are talking about a general statement that is supposed to work for all n, and I might switch it to k to for the step that X3 was just talking about, because, I might want to say, OK, I'm just talking about some arbitrary value. I'm not talking about all natural numbers any more. So, I'm not really picky about that, but some people would say what you would show is for all k, for any k, $8^k - 3^k$ implies $8^{k+1} - 3^{k+1}$...If one is divisible by 5, the other is

divisible by 5. That's what we need to do. So, do we have any ideas about how to do that? We've already got the first part, but can we show working for the kth case would imply working for $k + 1$th case, for any k? If we can do that, X_3 is saying we'd be done. Do you have an idea X_5?
X_5: I don't have an idea, but I do have a question.
N: Yeah.
X_5: The "for all" part, do we just kind of assume that ...um, just take that for granted?
[N1 52:00]

X_5's question indicates a tension between the quantification of the inductive implication and the language used to quantify the inductive assumption– a tension that the introduction of $n = k$ might resolve, through PUG. Namely, adopting k as a fixed but arbitrary value of n, assuming $P(n)$ for $n = k$ (not for all values of n), and proving that $P(k + 1)$ follows, is sufficient for proving the universally quantified implication. Instead of discussing this distinct role of k, the instructors had treated the introduction of k as a shift in notation when moving from the proposition (quantified by n) to the inductive implication (quantified by k). To students, the introduction of k seemed pedantic.

Conclusions

Our study was driven by a desire to understand the challenges of proof by MI that might persist, even in response to research-based instruction. Prior literature has promoted QI as a particularly promising approach to formal MI, but one that leaves a gap (Harel, 2001). In our own prior work, we investigated the gap through clinical interviews with individual students, using a Piagetian framework to test Dubinsky's (1991) conjecture about the role of logical implication (held as an action or object) in mastering proof by MI (Arnold & Norton, 2017; Norton & Arnold, 2017, 2019). The present study took a whole-class approach to investigate how students might experience and address the gap within a research-based instructional setting. Our findings provide clearer insights into the gap, why it persists, and how we might further modify instruction to address it.

Instruction on MI often highlights the role of recursion, using metaphors like climbing stairs or knocking down dominoes (Ron & Dreyfus, 2004), but in line with prior research, we contend that students readily engage in recursive reasoning even before instruction on proofs and proving (Smith, 2002). Although we did use metaphors in our classes (e.g., MI as "an engine"), our research-based instructional approach featured QI and inquiry about the roles of various components of MI (e.g., the base case and the inductive implication; see Table 1). In this context, students not only readily engaged in recursive reasoning, they also experienced an intellectual need to generalize QI arguments so that the logic "repeats itself."

Prior literature has attributed students' struggles in formalizing inductive arguments to a conflation between the inductive assumption and the proposition itself (e.g., Movshovitz-Hadar, 1993). Indeed, we found several cases where students appeared to do just that. However, this ostensible conflation is not something that students generally experience. Students do make distinctions between the proposition and the inductive assumption, and they attempt to articulate those distinctions in various ways. Therefore, students seem to experience the cognitive gap, not as a conflation, but rather as an ambiguity in language, especially related to quantifying the inductive implication and the inductive assumption. Here, we support and elaborate on this main finding with a discussion of the five themes that arose from our study. Then, we consider the collective implications of these themes for instructional practice.

Discussion of Themes

As illustrated in Fig. 4, the themes surrounding the apparent conflation between the inductive assumption and the original proposition emanate in two directions from the cognitive gap. In addressing the gap, students begin to establish an intellectual need for–and language that supports–a quantitative distinction between the inductive assumption and the assumption of the proposition. Ultimately, the intellectual need for this distinction should justify the special role that the new variable k plays in proof by MI.

When engaged in QI, students in our study experienced an intellectual need to generalize QI arguments so that those arguments would continue indefinitely (Theme 1). Buttressing this need, the instructors promoted MI as a labor-saving device, sometimes using the metaphor of an engine that would do work for them once started. While students accepted MI as a method that would satisfy their intellectual need, some students were also satisfied by less rigorous arguments. For example, students might rely on a kind of "method of descent" (cf. Ernest, 1982) through which an arbitrary nth case might be reduced to smaller and smaller cases through some unprescribed limiting process.

When students attempted formal MI, the issue of quantification arose quickly. Although many of their spoken and written quantifications of the inductive hypothesis ostensibly assumed the proposition, students held distinctions between the proposition and their assumptions (Theme 2). They sometimes used "if" and "such that" as tentative quantifiers, arguing that such terms restrict their assumptions to cases in which the inductive hypothesis is known to be true. This issue was especially prevalent as students attempted to shift quantification from the statement of the inductive implication to its proof. Whereas the former requires a universal quantifier, the latter should be assumed for a fixed but arbitrary case. The connection between these two quantifications relies on PUG (Copi, 1954; Dawkins & Roh, 2016), a principle rarely discussed in literature on proof by MI.

Students required new language to support the distinctions they were beginning to make between quantification of the proposition, inductive implication, and the inductive assumption (Theme 3). The issue of utilizing appropriate language to express quantification was further exacerbated by interpersonal ambiguity in use of terms like "for any" and "for some." This finding is remarkably similar to the finding of Lew and Mejía-Ramos' (2019) that students use quantifying terms unconventionally, which often leads to over/under-quantification. However, even with ambiguities resolved and conventional uses of terms established, one critical language-related issue remains–the aforementioned shift in quantifying the inductive implication and the inductive assumption used to prove it.

The role of k as a fixed but arbitrary quantity, distinct from n, was mentioned in both classrooms, but students did not demonstrate an intellectual need for it to support the language of quantification (Themes 4 & 5). In fact, Norton permitted his students to ignore the shift in notation (from n to k) altogether. Teachers did not highlight how k may be used to emphasize that in proving the inductive implication, the inductive hypothesis is assumed for a single value of n, not for all n. Furthermore, the need for k might have arisen if the instructors had initiated discussions of PUG and the dependence of MI on that principle, particularly in proving the inductive implication. Students typically rely on PUG when proving any "for all" statement, but a proof by MI does not follow this standard format. We elaborate on this distinction in the next section when we consider directions for future research.

Focusing on the Link Between PUG and MI in Future Research

Our study affirms the instructional value of approaching MI through QI, which seems to induce an intellectual need for generalizing and continuing arguments. Additionally, findings from our study promote productive classroom discussions of the components of MI (see Table 1), their quantification, and the meanings of related terms (see Fig. 7). The instructors adopted these approaches in their own classrooms, which helped them home in on the gap that remains. This gap persisted in the nebulous distinction between the quantification of the inductive implication and the quantification of its assumption when proving it. It is important to note that despite differences in instruction, students in both classes exhibited similar difficulties. Thus, beyond our own research-based instructional approaches, we recommend that discussions of MI also include explicit discussions of PUG. These discussions could be facilitated by the distinction Dawkins and Roh (2020a, b) made between the role and the value of a variable. Although n and k might carry the same values, they serve different roles in proof by MI.

Prior to instruction on MI, students become accustomed to proving universally quantified statements via PUG. When proving a statement of the form "for all $n \in \mathbb{N}$, $P(n)$ is true" without MI, they begin their proof with "let $n \in \mathbb{N}$ be arbitrary." Then, they give a general argument for why $P(n)$ is true for this arbitrary n. In particular, n is fixed throughout the entire proof. A proof by MI of the same statement, however, does not apply PUG to the variable n in the usual way that students might anticipate. Not only does the proof not begin with the language "let $n \in \mathbb{N}$ be arbitrary," but n actually varies throughout the proof. In the base case, we set $n = 1$. Then, utilizing PUG, we fix n to be an arbitrary value k to demonstrate the inductive implication holds for all n. The proof of the inductive implication via PUG ultimately sits within a larger proof (like a lemma) that depends on the principle of MI.

The variation of n throughout a proof by MI can often lead students to erroneously conclude that the corresponding metaphorical "ladder" climbed is infinitely long. MI actually demonstrates that given a ladder, any specified rung can be reached; the position reached is finite. For example, students may erroneously try to prove by MI that the intersection of infinitely many nested, nonempty sets is nonempty by inducting on the number of sets that are intersected. They contend that such a proof establishes the intersection of infinitely many sets as nonempty. However, they have only demonstrated that for any specified *finite* collection of nested sets, the intersection is nonempty.

Students' familiarity with proofs that utilize PUG suggests an alternative structure for writing proofs by MI. Metaphorically, this style would begin by specifying a ladder with a fixed, but arbitrary number of rungs n. Then, the proof would demonstrate that the top rung can be reached. Rather than utilizing the variable k only during the proof of the inductive implication, k would be used throughout to represent the intermediate rung position of the climber. In this way, n remains fixed and k plays a role analogous to an indexing variable of a sum, e.g., $\sum_{k=1}^{n} a_k$, stepping from the base case up to n. The alternative proof might be written as follows:

Proof Let $n \in \mathbb{N}$ be arbitrary.

Base case: Let $k = 1$. Demonstrate $P(1)$ is true.

Inductive Implication: Fix arbitrary $k \in \mathbb{Z}$ such that $1 \leq k < n$. Assume $P(k)$ is true. Show $P(k + 1)$.

Thus, $P(n)$ must be true. □

Ultimately, the fact that $P(n)$ is true for all $n \in \mathbb{N}$ follows as usual by PUG.

Summary

Our study built on prior literature by designing instruction informed by it, particularly the use of QI (Avital & Libeskind, 1978; Harel, 2001). Furthermore, we coded student responses based on factors identified in prior research that might influence students' understanding of MI (Baker, 1996; Davis et al, 2009; Dubinsky, 1986, 1990; Ernest, 1984; Harel, 2001; Movshovitz-Hadar, 1993; Palla et al., 2012; Ron & Dreyfus, 2004; Stylianides et al., 2007). Through that coding process and the themes that arose from it, we were able to identify more nuanced aspects of students' struggles. At the same time, we characterized factors that seem to play a lesser or different role, such as questions about students' abilities to engage in recursive reasoning and the inference that students conflate the proposition with the inductive assumption. In the end, our analysis of students' reasoning brought us to an unexpected conclusion regarding PUG (Copi, 1954). Although we had considered the critical role students' understanding of logical implication would play in the understanding of MI (Arnold & Norton, 2017; Norton & Arnold, 2017, 2019), and although prior literature has anticipated the role of PUG in other forms of logical reasoning (Dawkins & Roh, 2016), we had not considered the special role PUG would play (via k) in students' understanding of MI. Our study implies that instruction should explicitly address that role, through classroom discussion, to bridge the gap between QI and MI.

Data Availability Video data is protected by IRB but coded transcriptions are available upon request.

Declarations

Conflicts of Interest/Competing Interests The authors have no conflict of interest in the work reported here.

References

Arnold, R., & Norton, A. (2017). Mathematical actions, mathematical objects, and mathematical induction. In A. Weinberg, C. Rasmussen, J. Rabin, M. Wawro, and S. Brown (Eds.), *Proceedings of the 20th Annual Conference on Research in Undergraduate Mathematics Education,* (pp. 53-66). San Diego, California. (ISSN 2474-9346).

Avital, S., & Libeskind, S. (1978). Mathematical induction in the classroom: Didactical and mathematical issues. *Educational Studies in Mathematics, 9*(4), 429–438.

Baker, J. D. (1996). Students' difficulties with proof by mathematical induction. Paper presented at the *Annual Meeting of the American Educational Research Association* (April 8-12), New York. https://doi.org/10.3102/0013189X024008037

Copi, I. (1954). *Symbolic logic*. New York: The Macmillan Company.

Cusi, A., & Malara, N. A. (2008). Improving awareness about the meaning of the principle of mathematical induction. In O. Figueras, J. L. Cortina, S. Alatorre, R. Rojano, & A. Sepúlveda (Eds.), *Proceedings of the Joint Meeting of PME-32 and PME-NA 30* (Vol.2, pp. 393-398). Morelia, Mexico: Cinvestav-UMSNH.

Davis, M., Grassl, R., Hauk, S., Mendoza-Spencer, B., & Yestness, N. (2009). Learning proof by mathematical induction. In M. Zandieh (Ed.), *Proceedings of the 12th Conference on Research in Undergraduate Mathematics Education* (February 26 - March 1, 2009), Raleigh, North Carolina. http://sigmaa.maa.org/rume/crume2009/

Dawkins, P. C., & Roh, K. H. (2016). Promoting metalinguistic and metamathematical reasoning in proof-oriented mathematics courses: A method and a framework. *International Journal of Research in Undergraduate Mathematics Education, 2*(2), 197–222.

Dawkins, P. C., & Roh, K. H. (2020a). Assessing the influence of syntax, semantics, and pragmatics in student interpretation of multiply quantified statements in mathematics. *International Journal of Research in Undergraduate Mathematics Education, 6*(1), 1–22.

Dawkins, P. C., & Roh, K. H. (2020b). Coordinating two meanings of variables in proofs that apply definitions repeatedly. In Karunakaran, S. S., Reed, Z., & Higgins, A. (Eds), *Proceedings of the 23rd annual conference for Research in Undergraduate Mathematics Education,* (pp. 87-95).Boston, MA.

Dawkins, P. C., & Weber, K. (2017). Values and norms of proof for mathematicians and students. *Educational Studies in Mathematics, 95*(2), 123–142.

Dawkins, P. C., Zazkis, D., & Cook, J. P. (2020). How do transition to proof textbooks relate logic, proof techniques, and sets?. PRIMUS, 1–17.

Dawson, J. W. (2006). Why do mathematicians re-prove theorems? *Philosophia Mathematica, 14*(3), 269–286.

Dubinsky, E. (1986). Teaching mathematical induction: I. *The Journal of Mathematical Behavior, 5,* 305–317.

Dubinsky, E. (1990). Teaching mathematical induction: II. *The Journal of Mathematical Behavior, 8,* 285–304.

Dubinsky, E. (1991). Reflective abstraction in advanced mathematical thinking. In D. Tall (Ed.), *Advanced Mathematical Thinking* (pp. 95–123). Netherlands: Kluwer Academic Publishers.

Durand-Guerrier, V. (2003). Which notion of implication is the right one? From logical considerations to a didactic perspective. *Educational Studies in mathematics, 53*(1), 5–34.

Ernest, P. (1982). Mathematical induction: A recurring theme. *The Mathematical Gazette, 66*(436), 120–125.

Ernest, P. (1984). Mathematical induction: A pedagogical discussion. *Educational Studies in Mathematics, 15*(2), 173–189.

Fischbein, E., & Engel, I. (1989). Psychological difficulties in understanding the principle of mathematical induction. In G. Vergnaud, (Ed.), *Proceedings of the 13th International Conference for the Psychology of Mathematics Education* (pp. 276-282). Paris, France.

Hammock, R. (2018). *Book of proof* (3rd edition). http://www.people.vcu.edu/rhammack/BookOfProof/.

Harel, G. (2001). The development of mathematical induction as a proof scheme: A model for DNR-based instruction. In S. Campbell & R. Zaskis (Eds.), *Learning and Teaching Number Theory* (pp. 185–212). Norwood, New Jersey: Ablex Publishing Corporation.

Harel, G. (2008). What is mathematics? A pedagogical answer to a philosophical question. In R. B. Gold & R. Simmons (Eds.), Proof and other dilemmas: *Mathematics and philosophy*, (pp. 265–290). Washington, DC: Mathematical Association of America.

Knuth, E. J. (2002). Secondary school mathematics teachers' conceptions of proof. *Journal for Research in Mathematics Education, 33*, 379?405.

Knuth, E. J., Stephens, A. C., McNeil, N. M., & Alibali, M. W. (2006). Does understanding the equal sign matter? Evidence from solving equations. *Journal for Research in Mathematics Education, 37*(4), 297–312.

Lew, K., & Mejía-Ramos, J. P. (2019). Linguistic conventions of mathematical proof writing at the undergraduate level: Mathematicians' and students' perspectives. *Journal for Research in Mathematics Education, 50*(2), 121–155.

Michaelson, M. T. (2008). A literature review of pedagogical research on mathematical induction. *Australian Senior Mathematics Journal, 22*(2), 57–62.

Movshovitz-Hadar, N. (1993). The false coin problem, mathematical induction and knowledge fragility. *Journal of Mathematical Behavior, 12*(3), 253–268.

Norton, A., & Arnold, R. (2017). Logical implication as the object of mathematical induction. In E. Galindo & J. Newton, (Eds.), *Proceedings of the 39th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education (pp. 745-752).* Indianapolis, IN: Hoosier Association of Mathematics Teacher Educators.

Norton, A., & Arnold, R. (2019), Meeting the cognitive demands of proof by induction: the case of Ben. In S. Otten, A. G.. Candela, Z. de Araujo, C. Haines, & C. Munter (Eds.), *Proceedings of the forty-first annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. St Louis, MO: University of Missouri.

Palla, M., Potari, D., & Spyrou, P. (2012). Secondary school students' understanding of mathematical induction: structural characteristics and the process of proof construction. *International Journal of Science and Mathematics Education, 10,* 1023–1045.

Piaget, J. (1970). *Structuralism* (C. Maschler, Trans.). New York: Basic Books.

Ron, G., & Dreyfus, T. (2004). The use of models in teaching proof by mathematical induction. In M. J. Hoines & A. B. Fuglestad (Eds.), *Proceedings of the 28th Conference of the International Group for the Psychology of Mathematics Education, vol. 4* (pp. 113-120). Bergen, Norway.

Shipman, B. A. (2016). Subtleties of hidden quantifiers in implication. *Teaching Mathematics and its Applications, 35*(1), 41–49.

Smith, L. (2002). *Reasoning by mathematical induction in children's arithmetic*. New York: Pergamon.

Stylianides, G. J., Stylianides, A. J., & Philippou, G. N. (2007). Preservice teachers' knowledge of proof by mathematical induction. *Journal of Mathematics Teacher Education, 10*(3), 145–166.

Tsay, J., Hauk, S., Yestness, N. R., Davis, M., & Grassl, R. M. (2009). Teaching proof by mathematical induction: A preliminary report. In M. Zandieh (Ed.), *Proceedings of the 12th Conference on Research in Undergraduate Mathematics Education* (February 26 - March 1, 2009, Raleigh, North Carolina). Published electronically. http://sigmaa.maa.org/rume/crume2009/