PLASTICITY IN FIRE ANT QUEENS AND HUMAN TEACHERS:

CHARACTERIZING SUPERGENE-MEDIATED GENETIC ASSIMILATION

AND

UNDERGRADUATE TEACHER KNOWLEDGE DEVELOPMENT

by

ALEXANDER HILL WAUGH

(Under the Direction of Dr. Tessa C. Andrews and Dr. Brendan G. Hunt)

ABSTRACT

This dissertation consists of two distinct bodies of work. The first leverages a unique aspect of fire ant biology to investigate a potential role for a supergene in genetic assimilation. A supergene refers to a chromosome structural rearrangement, such as an inversion, that regulates the production of an alternative phenotype within a species. In fire ants, queens that carry the ancestral, non-inverted arrangement of the relevant chromosome exhibit plastic, environmentally induced phenotypes that resemble the phenotypes regulated genetically by the derived, inverted supergene. By comparing gene expression patterns associated with plasticity that associated with supergene-related genetic differences, I found evidence consistent with an evolutionary scenario where, via genetic assimilation, the fire ant supergene genetically fixed the production a phenotype ancestrally produced only as a plastic response to a certain set of environmental stimuli. This finding contributed to important goals in evolutionary biology related to unraveling the molecular basis of plastic traits, elucidating the genetic processes that sustain alternative phenotypes within species, and characterizing an evolutionary role of inversion-derived supergenes as a genomic architecture primed to facilitate genetic assimilation. The second body of work aims to improve undergraduate STEM education by characterizing variation and development of the knowledge teachers need to implement evidence-based instructional strategies in large enrollment classrooms. Over the course of several semesters, I interviewed early-career undergraduate STEM instructors before and after they taught the same focal lesson. Qualitative analysis of these interviews revealed variation in teachers' knowledge of what students think about specific topics, how to teach those specific topics, and their general knowledge of how people learn. Longitudinal comparison of these interviews afforded insights into the patterns of knowledge development for these instructors and the instructional practice changes associated with knowledge development. Insights from this body of work can inform teaching professional development that aims to put early-career STEM instructors on the fast track to effective implementations of evidence-based teaching strategies.

INDEX WORDS: Fire ants, Supergene, Phenotypic plasticity, Polymorphism, Genetic assimilation, RNA-sequencing, STEM education, Discipline-based education research, Undergraduate Education, Teacher knowledge, Pedagogical content knowledge, Pedagogical knowledge

PLASTICITY IN FIRE ANT QUEENS AND HUMAN TEACHERS: CHARACTERIZING SUPERGENE-MEDIATED GENETIC ASSIMILATION AND UNDERGRADUATE TEACHER KNOWLEDGE DEVELOPMENT

by

ALEXANDER HILL WAUGH

B.S., University of Georgia, 2019

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2024

PLASTICITY IN FIRE ANT QUEENS AND HUMAN TEACHERS:

CHARACTERIZING SUPERGENE-MEDIATED GENETIC ASSIMILATION AND

UNDERGRADUATE TEACHER KNOWLEDGE DEVELOPMENT

by

ALEXANDER HILL WAUGH

Major Professors: Tessa C. Andrews

Brendan G. Hunt

Committee: Paula P. Lemons

Michael A. White

Kenneth G. Ross

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
December 2024

DEDICATION

For my parents, Susan and Robert Waugh,

who taught me to search for fulfillment in life's cohesive moments of

motivation, intentionality, and spontaneity.

"The mind that is not burdened by the past,

that is not caught in the present,

and that is not projecting into the future,

is in a state of learning."

- Jiddu Krishnamurti

ACKNOWLEDGEMENTS

I owe an enormous debt of gratitude to my research mentors, Tessa Andrews and Brendan Hunt. They guided me when I needed it, showed patience when I made mistakes, and gave me room to become my own scientist. Their laboratories should be studied as exemplars of intellectually stimulating, emotionally supportive, methodologically rigorous, and genuinely enjoyable scientific environments. I would also like to thank the other members of my committee, Paula Lemons, Mike White, and Ken Ross, who invested in me and my career as a teacher and scientist by supporting my unusual dissertation work in two distinct fields.

To my lab mates, Hannah Ericson, Cristie Donham, Sandhya Krishnan, Gretchen King, Katie Green, Lexie Cooper, Benjamin Jackson, Samuel Arsenault, Horace Zeng, Sabrina Barbosa, and Alekz Cushman, I want to thank each of you for your feedback and support but, more importantly, for making the past 5+ years such an enjoyable experience. I need to give a special shout out to my friends C.J. Zajic, Benjamin Phipps, and Yitang Sun for their unending positivity, movie recommendations, and senses of humor. I also want to thank the larger Biology Education Research Group at UGA and the “Fire Ant Supergroup” who provided feedback on my work several times and fostered discussions that led me down important paths of thinking in my development as an education researcher and evolutionary biologist.

I love my family. My entire family has been so supportive over the course of my graduate school journey. My parents, Robert and Susan, instilled in me an insatiable curiosity and a strangely deep reservoir of confidence. They’ve given me everything a person needs to succeed. My brother, James, asked me questions that genuinely changed my thinking about my own research and reminded me of what interested me about my work on the rare occasions when morale was low. My sisters, Lucy and Emmy, are each a unique source of light in my life. They put up with my unsolicited elder sibling advice and do a fantastic job of being exactly who they are. I thank them for their unwavering support.

Lastly, I must acknowledge all my former teachers who played their part in encouraging, challenging, and inspiring me to take charge of my own education. I’m grateful for every minute I’ve had the privilege of spending in the classroom, and I aspire to have the same impact on students that you’ve had on me. Thank you. With that, I turn to you, the reader, and wish you happy trails on what promises to be a long but hopefully rewarding hike through this dissertation.

TABLE OF CONTENTS

Page

CHAPTER ONE

INTRODUCTION

A reader's guide to this dissertation

This dissertation includes quantitative research focused on the evolutionary genetics of fire ants and qualitative research on teacher knowledge for effective, evidence-based instruction at the undergraduate level. Fire ant queens exhibit phenotypic plasticity by the strict biological definition of this term. This plasticity served as the focus of the first body of work. Human teachers exhibit plasticity in the colloquial sense. Instructors develop new knowledge of how to teach and this development corresponds with changes to their teaching practices. Employing both the biological and colloquial definitions of the term, this dissertation focuses on plasticity in fire ant queens and human teachers.

In a dissertation that contains three research articles across two distinct academic fields, the reader benefits from a brief overview of this dissertation's structure and organization. The present chapter provides a brief review of the relevant aspects of fire ant genetics and the theory of genetic assimilation along with a discussion of the motivations for my discipline-based education research. Chapter 2 is a quantitative research article that characterizes the molecular underpinnings of phenotypic plasticity and explores the potential role of a chromosomal inversion acting as a mediator of increased genetic regulation of ancestrally plastic traits. Chapters 3 and 4 are discipline-based education research articles focused on qualitatively characterizing the variation and development of teacher knowledge for active learning in large undergraduate STEM courses. These articles emerged from analysis of different components of the same data set. Chapter 3 focuses on variation and development of pedagogical content knowledge, whereas Chapter 4 focuses on pedagogical knowledge of how people learn. In Chapter 5, I conclude with a discussion of opportunities for future research to build on these bodies of work.

Background Information: Supergene evolutionary biology and fire ants

Supergene evolutionary biology

For decades, fire ants have provided biologists a riveting front-row seat at the intersection of complex social behavior, phenotypic plasticity, and the evolutionary dynamics of large structural rearrangements of the genome also known as “supergenes”. Similar to the evolution of heteromorphic sex chromosomes from their autosomal origins in many systems, chromosomal inversions facilitate supergene evolution via suppression of meiotic recombination across non-collinear genomic arrangements (Bachtrog, 2013; Schwander et al., 2014). As a result, many genes and their potentially locally adapted alleles become tightly linked inside the non-recombining region. In the cases where such inversions are non-lethal, the inversion can cause the stable transmission of co-adapted alleles together as a single genomic module that collectively regulates a suite of phenotypes. Supergenes are so named because, despite often containing hundreds of genes and spanning large portions of chromosomes, linkage disequilibrium causes an entire chromosomal region to display a pattern of simple Mendelian inheritance typical of individual genes. Recombination suppression allows the maintenance within species of diverging genomic regions over evolutionary time despite admixture between carriers of those diverging regions.

However, supergenes also impose fitness costs mainly because of mutational disruptions to coding genes at inversion breakpoints, changes to regulatory dynamics that depend on the position of a gene on the chromosome, and the sequence divergence that ensues between the ancestral and derived arrangements (Charlesworth, 2016). This sequence divergence occurs because recombination suppression interferes with the capacity of selection to act on individual mutations. This phenomenon known as Hill-Robertson interference (Hill & Robertson, 1966) limits both the spread of advantageous mutations and the removal of deleterious ones, as natural selection is influenced by the surrounding genetic context. In other words, selection cannot act on advantageous and deleterious loci independently when they are in tight linkage. Consequently, the efficiency of selection is reduced, often leading to sequence degeneration in the non-recombining region. This degeneration typically manifests as the accumulation of fixed deleterious mutations, reduced genetic diversity, alterations to gene expression, gene loss, and the proliferation of transposable elements (e.g*.,* Bachtrog et al., 2011; Stolle et al., 2019; Tuttle et al., 2016).

The fire ant supergene

In S. invicta, a chromosome structural rearrangement consisting of a series of at least three inversions spanning more than 11 Mb (~50%) of chromosome 16 governs the social form of the colony (Ross & Keller 1998; Wang et al., 2013; Pracana et al., 2017; Yan et al., 2020). This supergene contains more than 500 of the roughly 16,000 protein-coding genes in the *S. invicta* genome (Yan et al., 2020). Among these genes is Gp-9, an odorant-binding protein with two allelic variants (Gp-9^B and Gp-9^b). The strong association of Gp-9 with social form in *S. invicta* led to the discovery of an inversion-based supergene and its regulation of fire ant social organization (Ross & Keller 1998; Wang et al., 2013). The “social b” (*Sb*) supergene haplotype contains the Gp-9^b allele and rarely recombines with the non-inverted “social B” (*SB*) ancestral haplotype which contains Gp-9^B (Pracana et al., 2017; Ross & Shoemaker, 2018; Yan et al., 2020). In monogyne colonies, the inferred ancestral social form (Ross & Keller, 1995), the queen is always homozygous (*SB/SB*) for the ancestral, non-inverted arrangement. Polygyne reproductive queens are nearly always heterozygous (*SB/Sb*) for the derived arrangement in U.S. populations (Ross, 1997).

Over the course of the estimated 0.5M year evolutionary history of the fire ant supergene (Pracana et al., 2017), two core factors have combined to drive sequence divergence between *Sb* and *SB*. As a result of their sequence non-collinearity, *SB* and *Sb* rarely recombine (Pracana et al., 2017; Ross & Shoemaker 2018; Yan et al., 2020), resulting in the accumulation of mutations, transposable elements, and other deleterious sequences in Sb since selection efficiency is reduced in this region (Huang et al., 2018; Stolle et al., 2019; Yan et al., 2020). One might expect meiotic recombination between *Sb* haplotypes in *Sb/Sb* homozygotes to rescue at least some of the lost selection efficiency in the derived arrangement. However, at the population level, recombination between *Sb* haplotypes is effectively suppressed by the low fitness of *Sb/Sb* individuals. *Sb/Sb* queens, and to a lesser extent *Sb* males, are unfit and rarely survive to mate as an apparent consequence of harboring only the inverted arrangement (Hallar et al., 2007). Surveys of U.S. polygyne populations estimate *Sb/Sb* queens make up less than 1% of polygyne reproductive queens in S. invicta (Ross, 1997; DeHeer et al., 1999).

The remainder of the fire ant genome outside of the non-collinear *Sb* region is freely recombining. *Sb* represents the only major structural rearrangement and the only large region of elevated linkage disequilibrium or sequence divergence in the S. invicta genome (Pracana et al., 2017; Yan et al., 2020). Thus, two core factors have driven substantial but localized sequence divergence between *Sb* and *SB*: (1) infrequent recombination between *Sb* and *SB* and (2) a low population-level recombination rate between *Sb* haplotypes. Since their discovery, the SB and Sb haplotypes have attracted interest from various subfields of biology, establishing *S. invicta* as an emerging model system for studying supergenes and complex phenotypes. In my research, I focus on how Sb regulates polymorphisms of weight and colony-founding behavior between *SB/SB* and *SB/Sb* gynes (Keller & Ross 1993a, 1993b, 1998, 1999; Ross & Keller 1998; DeHeer, 2002).

Fire ant gyne weight and colony-founding behavior

Ants exhibit two modes of colony-founding known as independent and dependent colony-founding. Each mode is fundamentally differentiated by the founding queen's dependence on worker assistance in rearing her initial offspring (Hölldobler & Wilson, 1977). Independently founding queens start their colony from scratch and use their own nutrient reserves to rear their initial brood. Dependently founding queens are unable to rear their initial brood without worker assistance, meaning they must initiate egg-laying in a pre-existing nest. A gyne's supergene genotype is generally predictive of her weight and colony-founding strategy (DeHeer et al., 1999). Both monogyne and polygyne *SB/SB* gynes found colonies independently whereas *Sb*-carrying gynes found colonies dependently. A queen's weight also shows strong positive correlation with her fecundity. The lighter and less fat *SB/Sb* gynes have less developed ovaries at sexual maturity, and reach sexual maturity slower compared to *SB/SB* gynes of either social form (Vargo & Fletcher 1989; Keller & Ross 1993a, 1993b; Keller & Ross 1999, DeHeer, 2002).

The weight polymorphism between *SB/SB* and *SB/Sb* gynes is influenced by the different selective pressures associated with their distinct colony-founding modes (Hölldobler & Wilson, 1977; Keller & Passera, 1989). The more nutrient-rich *SB/SB* queens have an advantage of greater dispersal distances during their mating flight and subsequently found new colonies independently using primarily their substantial fat reserves to feed their initial brood (Keller & Passera, 1989). These initial brood will develop into workers which permanently take over the responsibility of feeding new brood as the colony begins to grow. The dispersal capability and fecundity of *SB/SB* queens is crucial to the survival of an incipient colony. Longer dispersal allows independently founding queens to reach areas with more resources and less competition. Fecundity provides a distinct advantage during independent colony-founding as queens that can quickly rear the largest initial workforce have an advantage during the competition of independent colony-founding.

Compared to *SB/SB* gynes of either social form, polygyne *SB/Sb* gynes are lightweight and nutrient-poor and lack the metabolic reserves for long mating flights and independent colony-founding (DeHeer et al., 1999; DeHeer, 2002). These deficiencies can likely be attributed to the sequence degeneration and regulatory disruptions associated with the inverted *Sb* haplotype. Instead of attempting independent founding, *SB/Sb* gynes found colonies dependently by invading a nearby polygyne nest (DeHeer et al., 1999*)*. Polygyne workers accept these invading *Sb*-carrying gynes as new queens and take over the responsibility of feeding their brood. As a now classic example of a “green beard” phenotype, polygyne workers accept additional queens so long as each queen bears a pheromonal signature associated with *Sb* (Dawkins, 1976; Ross & Keller 1998; Trible & Ross 2016). A “green beard” refers to a heritable trait that allows individuals to recognize and preferentially help others who carry the same trait, promoting cooperative behavior. Short flights, *Sb*-mediated queen acceptance, and dependent founding create conditions where polygyne colonies can afford to reduce their resource investment in producing heavier gynes and lightweight *Sb*-carrying gynes can still reproduce despite limited nutrient reserves.

To summarize the background information provided so far, *SB/SB* gynes produced by either monogyne or polygyne colonies are heavy, bearing robust nutrient reserves and capable of rapid oogenesis. These heavyweight phenotypes are crucial for feeding an initial brood during independent founding. *Sb*-carrying gynes are lighter, bearing relatively fewer nutrient reserves and exhibiting slower oogenesis. Typically, an ant species will specialize in just one of the two colony-founding modes. However, the fire ant displays two genetically regulated forms of social organization, monogyny and polygyny regulated by a supergene, with supergene genotype acting as a near perfect predictor of gyne weight and colony-founding mode.

Plasticity in fire ant gynes associated with overwintering

Phenotypic plasticity refers to the capacity of a single genotype to produce a range of phenotypes in response to different environmental conditions (Stearns, 1989; West-Eberhard, 2003). Whereas evolution occurs by changing allele frequencies in a population over many generations, phenotypic plasticity results in phenotypic change within an organism's lifetime as a response to an environmental stimulus. Plasticity gives rise to variation in morphology, physiology, and behavior across the tree of life. The bar-tailed godwit (*Limosa lapponica*) provides an example of each of these forms of phenotypic plasticity occurring within a single organism. This shorebird seasonally alters its body composition, increasing fat stores and muscle mass and decreasing the size of their nutritional organs before embarking on long-distance, non-stop flights (Piersma, 1998). In fact, the bar-tailed godwit holds the record for the longest recorded non-stop flight at 8,245 miles over a period of 11 days between Alaska and Australia (Leffer, 2021). Plasticity enables them to store sufficient energy for migration and reduce the metabolic cost of maintaining organs non-essential for flight. Remarkably, this bird also behaviorally responds to environmental cues including barometric pressure to decide when to begin their flight and how to adjust along the way (Gill et al., 2014).

A recently discovered life history strategy of overwintered monogyne fire ant gynes (*SB/SB*), demonstrates environmentally induced plasticity in gyne weight and colony founding behavior. Certain monogyne *SB/SB* gynes that eclose late in the mating season overwinter in their natal nest and proceed to lose weight over the course of the winter (Tschinkel and Howard, 1978; Tschinkel 1996; DeHeer & Tschinkel, 1998; Fletcher & Blum 1983; Helms & Godfrey 2016). When they emerge to take their postponed flight the following spring, overwintered gynes are much lighter than their spring-reared, non-overwintered *SB/SB* counterparts, but they do not differ in other aspects of morphology such as head width, thorax size, etc. This suggests that overwintered and spring-reared gynes experience similar developmental programs, and that the weight reduction evident among overwintered gynes is a plastic response to an environmental stimulus. Fire ants display both a supergene-mediated polymorphism affecting weight and colony founding behavior and overwintering-related plasticity in these traits. After overwintering, lightweight, dependently founding *SB/SB* gynes in monogyne nests provide the exception that renders supergene genotype only a near perfect predictor of gyne weight and colony-founding behavior.

Genetic assimilation and the fire ant supergene

Genetic assimilation is the evolutionary process by which a phenotype that initially arises as a response to an environmental condition becomes increasingly genetically regulated or fixed over time (Waddington, 1953; West-Eberhard, 2003; Pigliucci et al., 2006; Moczek et al., 2011; Pfennig & Ehrenreich, 2014; Ehrenreich & Pfennig, 2015; Jones & Robinson, 2018; Nijhout et al., 2021; Wood et al., 2023). Genetic assimilation occurs when a trait is initially produced as a response to an environmental stimulus, without any genetic change. Then, the environmentally induced trait becomes evolutionarily advantageous, and natural selection favors individuals who express the trait under certain conditions. With ongoing selection, the trait becomes more frequent in the population. Eventually, mutations accumulate, causing organisms to develop the trait even in the absence of the original environmental stimulus. Ultimately, via genetic assimilation, an ancestrally plastic trait becomes a genetically fixed in the population, no longer requiring environmental triggers for its expression.

As an example of genetic assimilation, *Spea* spadefoot toad tadpoles display two distinct “ecomorphs”, each with unique morphological and behavioral traits. The omnivore ecomorph is the ancestral, default ecomorph. When these tadpoles eat shrimp or other tadpoles, a carnivorous ecomorph is induced via phenotypic plasticity. When only one species of *Spea* is present in a population, frequency-dependent, disruptive selection maintains both ecomorphs, likely due to selective pressures arising from limited food resources. When two species of *Spea* co-occur, one species specializes in the production of the omnivorous and the other the carnivorous ecomorph. This specialization occurs by the loss of ancestral plasticity and the increased genetic regulation of carnivore phenotypes, as demonstrated by the production of such traits in this population irrespective of environmental cues (Pfennig & Murphy, 2000).

Chromosomal inversion-derived supergenes provide both immediate and long-term opportunities for selection to shape the genetic mechanisms controlling phenotypic plasticity. Genetic assimilation can occur via regulatory and coding sequence evolution since either has the potential to alter or disrupt molecular machinery underlying plasticity to buffer against environmental variation (Scoville & Pfrender, 2010; Ehrenreich & Pfennig, 2015; Levis et al., 2017; Wood et al., 2023). Inversions suppress meiotic crossover events and thus can promote the accumulation and fixation of mutations that affect copy numbers and sequences of non-coding, regulatory, and protein coding elements (Hill & Robertson, 1966; Feder et al., 2011; Bachtrog, 2013; Pracana et al., 2017; Wellenreuther & Bernatchez, 2018; Faria et al., 2019; Stolle et al., 2019; Fontana et al., 2020; Martinez-Ruiz et al., 2020). These mutations can disrupt the molecular machinery underlying the plastic phenotype, resulting in supergene-mediated genetic assimilation. Supergenes may also drive genetic assimilation by suppressing crossover events during meiotic recombination, leading to sequence degeneration. These mutations can become fixed and maintained in tightly-linked allelic combinations (Wellenreuther & Bernatchez, 2018). This can stabilize phenotypes, even when environmental conditions fluctuate, by limiting plasticity through genetic regulation.

In Chapter 2, I examine this hypothesis by assessing whether genes associated with plasticity in gyne nutrient reserves have been perturbed in the course of supergene evolution. Specifically, I compared gene expression patterns associated with the supergene's regulation of polymorphic gyne weight and colony-founding behavior to the patterns of gene expression associated with plasticity in *SB/SB* gynes. This plasticity in the ancestral monogyne form yields *SB/SB* gyne phenotypes that seemingly recapitulate the lightweight and dependent colony-founding phenotypes typical of *Sb*-carrying gynes.

Motivating the work: DBER focused on teacher knowledge

Overview of DBER and my work in the field

Discipline-Based Education Research (DBER) is a field of research focused on education within specific academic disciplines, such as biology, chemistry, physics, engineering, or mathematics. DBER investigates teaching and learning with the goal of improving educational practices in these disciplines by integrating knowledge from education research with disciplinary knowledge of the concepts, practices, and educational challenges specific to a given academic field. As a relatively new field, DBER is expanding to encompass a broad range of subfields, including critical research aimed at addressing the educational debt owed to marginalized members of our communities. However, much of DBER focuses on arriving at an empirical understanding of how students learn in specific disciplines as well as the development and assessment of teaching methods that enhance student learning outcomes.

With crystal clarity, DBER has coalesced around central findings regarding the principles of how students learn and the teaching strategies that best capitalize on these principles. Students learn best by engaging in active learning. Active-learning instruction encompasses diverse teaching methods, defined by the observable behaviors students demonstrate as they engage deeply in conceptual learning. These behaviors involve students going beyond what has been presented to them in class to generate representations of their thinking such as explanations of their reasoning in dialogue with peers, written experimental designs, or schematic diagrams that model scientific phenomena (Chi & Wylie, 2014).

Time and time again, empirical research shows that active-learning instruction has the potential to increase students' abilities to learn fundamental concepts (e.g., Freeman et al., 2014) while beginning to repay the educational debt owed to minoritized students (e.g., Theobald et al., 2020). A widely cited meta-analysis of 225 studies in the *Proceedings of the National Academy of Sciences* compared examination scores and failure rates in "treatment" courses that used active learning versus “control” courses that featured only traditional lecture (Freeman et al., 2014). The benefits of active-learning proved so substantial that if the studies “had been conducted as randomized controlled trials of medical interventions, they may have been stopped for benefit—meaning that enrolling patients in the control condition might be discontinued because the treatment being tested was clearly more beneficial” (Freeman et al., 2014). It’s also worth noting that leading educational psychologists called for a ban on traditional lecture in classrooms nationwide over two decades ago (Crouch & Mazur, 2001).

Despite overwhelming evidence of the benefits of active-learning compared to traditional, “teaching by telling” lecture, individual instructors encounter substantial challenges when they begin implementing active-learning instruction in their classrooms. These challenges become evident as groups of instructors implement similar active-learning instructional strategies but achieve disparate student learning outcomes (Andrews et al., 2011; Laursen et al., 2014; Johnson et al., 2020). So, what gives? If evidence for the learning benefits of active-learning instruction is so overwhelming, why do many instructors struggle to achieve the benefits promised in the literature using the prescribed instructional strategies? The answer lies in the nuances of their implementation when they adapt active-learning instructional strategies to their classrooms.

The nuances in how an instructor implements an evidence-based, active-learning instructional strategy matter. For instance, instructors could use group work with the intention of fostering dialogue between students which should promote learning. However, their adaptation of this active-learning strategy could then fail because the instructor explained the content to students as they interacted with student groups instead of leaving them the opportunity to learn by going beyond what has been presented to them and constructing their own knowledge. Additionally, despite the benefits of interacting with students one-on-one during group work touted in the literature, instructors who lack knowledge of how students tend to think about a particular topic may struggle to anticipate and respond effectively to a student’s reasoning as they interact with them during class. Over the course of a semester, the lack of fidelity between an instructor's implementation of an active-learning strategy and its implementation as studied in the literature yields the discrepancy often observed between hypothesized and realized student learning outcomes (e.g., Andrews et al., 2011; Dancy et al., 2016; Nehm et al., 2022).

My research aims to address this discrepancy. I approach my research by acknowledging that decades of work have produced a robust set of evidence-based, active-learning instructional strategies that have the potential to benefit students. Fundamentally, I aim to complement this work by characterizing what I see as a missing piece: the knowledge instructors need to effectively adapt active-learning instructional strategies to their classrooms. In the body of work presented in Chapters 3 and 4, I characterize the variation of this knowledge among early-career active-learning instructors teaching in large-enrollment undergraduate classrooms. I also describe the patterns of its development and how this knowledge development corresponds with changes in teaching practices. Ultimately, the goal of this work is to help instructors develop the expertise they need to use active learning effectively in the modern undergraduate classroom.

Teacher knowledge for effective active-learning instruction

What knowledge differentiates the leading expert in academic research on photosynthesis and the leading expert on teaching photosynthesis to undergraduate students in an introductory biology course? Assuming these experts are different individuals, we might predict the academic researcher possesses deeper content knowledge of photosynthesis since the researcher stays up to date reading the most current literature and the instructor teaches photosynthesis as just one of many topics in their course. We also might predict that, even if they could teach their preferred lesson topic of photosynthesis, the researcher will struggle to facilitate learning effectively if they lack the knowledge of the expert instructor. *However, what does this expert instructor know that the researcher does not? And how does this knowledge* *relate to what they do in the classroom?* Two components of teacher knowledge demonstrated to reliably differentiate these two individuals served as the foci of my qualitative education research: pedagogical content knowledge and pedagogical knowledge of how people learn.

Pedagogical content knowledge (PCK) refers to an instructor's awareness of how students think about specific topics within a discipline and how to teach these topics (Gess-Newsome, 2015). PCK enables instructors to anticipate, access, and respond to patterns of student thinking as instructors plan, teach, and reflect on a lesson (Gehrtz et al., 2022). By its very nature, active learning involves students going beyond what is presented, which gives the task of the instructor a quasi-improvisational element. This makes it essential for instructors to be aware of how students are likely to respond to their instructional strategies and to know how to intervene effectively to support learning when these responses occur. PCK is especially beneficial in the active learning environment compared to the traditional lecture hall. Whereas the traditional lecture involves unidirectional information flow from teacher to student, information flow in an active learning classroom is dynamic and bi-directional. Information flows from teacher to student, from student to teacher, and then back to the student. In this environment, PCK provides the instructor with the distinct advantage of knowing what difficulties are likely to arise and the appropriate instructional strategies to deploy when they do.

Pedagogical knowledge of how people learn is a category of teacher knowledge distinct from PCK that encompasses ideas and reasoning about the experiences that foster learning for students and how instructors can create these experiences. Whereas PCK involves knowledge of how students think about and learn specific topics, pedagogical knowledge of how people learn refers to knowledge of how learning works more generally and how to align teaching practices with those general learning principles. To begin thinking about the knowledge of learning instructors need, it helps to begin with an empirical understanding of how learning works. Given the complexity of learning as a cognitive process, researchers benefit from focusing on observable student behaviors to identify those that predict better or worse learning outcomes. The ICAP framework (Chi & Wylie, 2014) finds broad empirical support in identifying four discrete categories of cognitive engagement defined by specific classroom student learning behaviors: interactive (e.g., defending and arguing a position), constructive (e.g., predict the effect of mutations in a pathway), active (e.g., taking verbatim notes), passive (e.g., listening). Students learn more as they move up this hierarchy of behaviors, and they learn the most when learning is interactive (i.e., I>C>A>P). Pedagogical knowledge of how people learn leverages this concrete framework of student learning behaviors as it describes teacher knowledge, not of the ICAP framework itself but of its underlying learning principles and allied teaching practices. Theory predicts that pedagogical knowledge of how people learn gives the instructor the advantage of understanding the general learning principles as they design, adapt, and facilitate opportunities for learning in their classroom (Auerbach & Andrews, 2018; Auerbach et al., 2018; Andrews et al., 2019).

The advantages of PCK and pedagogical knowledge of how people learn may now seem obvious. If an instructor wants to facilitate learning, then shouldn't they be formally trained in how students tend to think about specific topics and how learning works in general? The importance of these components of teacher knowledge are demonstrated at the K12 level (e.g., Park et al., 2011), and early studies suggest they are also important at the undergraduate level (e.g., Andrews et al., 2019). However, this evidence appears only to inform the model for K12 teacher preparation. We train K12 teachers in the subject matter they're going to teach, the ways students approach thinking about specific topics, and the fundamental principles of student learning they will need as they enter their positions. However, at the undergraduate level, this formal training is not required of many instructional positions, and teaching professional development is often limited in availability and scope (e.g., Schussler et al., 2015). So, when do we give undergraduate instructors a chance to develop the teacher knowledge they need? The answer is that, by and large, we don't, and we make the often-fallacious assumption that these instructors will develop this knowledge with teaching experience (e.g., Chan & Yung, 2018).

Conservatively, around half of U.S. undergraduate STEM classrooms utilize active learning in some capacity (Freeman et al., 2014; Stains et al., 2018). On one hand, wide-scale adoption is a sign of progress. However, considering the challenges instructors face with active learning instruction and the lack of teaching professional development, it also poses an important problem for the higher education community to solve. Currently, nationwide efforts aim to make opportunities for teaching professional development more available, systematized, and impactful so we can better support instructors at the undergraduate level. The second body of work presented in this dissertation contributes to these efforts by characterizing the variation and development of PCK (Chapter 3) and pedagogical knowledge of how people learn (Chapter 4).

References

Andrews, T. M., Leonard, M. J., Colgrove, C. A., & Kalinowski, S. T. (2011). Active learning not associated with student learning in a random sample of college biology courses. *CBE—Life Sciences Education*, *10*(4), 394-405.

Andrews, T. C., Auerbach, A. J. J., & Grant, E. F. (2019). Exploring the relationship between teacher knowledge and active-learning implementation in large college biology courses. *CBE—Life Sciences Education*, *18*(4), ar48.

Auerbach, A. J. J., & Andrews, T. C. (2018). Pedagogical knowledge for active-learning instruction in large undergraduate biology courses: a large-scale qualitative investigation of instructor thinking. *International Journal of STEM Education*, *5*, 1-25.

Auerbach, A. J., Higgins, M., Brickman, P., & Andrews, T. C. (2018). Teacher knowledge for active-learning instruction: Expert–novice comparison reveals differences. *CBE—Life Sciences Education*, *17*(1), ar12.

Bachtrog, D., Kirkpatrick, M., Mank, J. E., McDaniel, S. F., Pires, J. C., Rice, W., & Valenzuela, N. (2011). Are all sex chromosomes created equal?. *Trends in genetics : TIG*, *27*(9), 350–357. https://doi.org/10.1016/j.tig.2011.05.005

Bachtrog, D. (2013). Y-chromosome evolution: Emerging insights into processes of Y-chromosome degeneration. *Nature Reviews Genetics*, *14*(2), 113–124.

Chan, K. K. H., & Yung, B. H. W. (2018). Developing pedagogical content knowledge for teaching a new topic: More than teaching experience and subject matter knowledge. *Research in Science Education*, *48*, 233-265.

Charlesworth, D. (2016). The status of supergenes in the 21st century: recombination suppression in Batesian mimicry and sex chromosomes and other complex adaptations. *Evolutionary Applications*, *9*(1), 74-90.

Chi, M. T., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational psychologist*, *49*(4), 219-243.

Crouch, C. H., & Mazur, E. (2001). Peer instruction: Ten years of experience and results. *American journal of physics*, *69*(9), 970-977.

Dancy, M., Henderson, C., & Turpen, C. (2016). How faculty learn about and implement research-based instructional strategies: The case of peer instruction. *Physical Review Physics Education Research*, *12*(1), 010110.

Dawkins, R. (1976). The Selfish Gene. Oxford University Press, New York

DeHeer, C. J. (2002). A comparison of the colony-founding potential of queens from single- and multiple-queen colonies of the fire ant Solenopsis invicta. *Animal Behaviour*, *64*(4), 655-661.

DeHeer, C. J., Goodisman, M. A. D., & Ross, K. G. (1999). Queen Dispersal Strategies in the Multiple-Queen Form of the Fire Ant Solenopsis invicta *153*(6), 660-675.

DeHeer, C. J., & Tscbinkel, W. R. (1998). The success of alternative reproductive tactics in monogyne populations of the ant Solenopsis invicta: significance for transitions in social organization. *Behavioral Ecology*, *9*(2), 130-135.

Ehrenreich, & Pfennig. (2015). Genetic assimilation: a review of its potential proximate causes and evolutionary consequences. *Annals of Botany, 117*(5), 769-779. doi:10.1093/aob/mcv130

Feder, J. L., Gejji, R., Powell, T. H., & Nosil, P. (2011). Adaptive chromosomal divergence driven by mixed geographic mode of evolution. *Evolution*, *65*(8), 2157-2170.

Fletcher, D. J., & Blum, M. S. (1983). The inhibitory pheromone of queen fire ants: effects of disinhibition on dealation and oviposition by virgin queens. *Journal of comparative physiology*, *153*, 467-475.

Fontana, S., Chang, N. C., Chang, T., Lee, C. C., Dang, V. D., & Wang, J. (2020). The fire ant social supergene is characterized by extensive gene and transposable element copy number variation. *Molecular Ecology*, *29*(1), 105-120.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the national academy of sciences*, *111*(23), 8410-8415.

Gehrtz, J., Brantner, M., & Andrews, T. C. (2022). How are undergraduate STEM instructors leveraging student thinking?. *International Journal of STEM Education*, *9*(1), 18.

Gess-Newsome, J. (2015). A model of teacher professional knowledge and skill including PCK: Results of the thinking from the PCK summit. In J. L. A. Berry & P. Friedrichsen (Eds.), *Reexamining pedagogical content knowledge in science education* (pp. 28–42). Routledge.

Gill Jr, R. E., Douglas, D. C., Handel, C. M., Tibbitts, T. L., Hufford, G., & Piersma, T. (2014). Hemispheric-scale wind selection facilitates bar-tailed godwit circum-migration of the Pacific. *Animal Behaviour*, *90*, 117-130.

Hallar, B. L., Krieger, M. J. B., & Ross, K. G. (2007). Potential cause of lethality of an allele implicated in social evolution in fire ants. *Genetica*, *131*(1), 69–79.

Helms IV, J. A., & Godfrey, A. (2016). Dispersal polymorphisms in invasive fire ants. *PLoS One*, *11*(4), e0153955.

Hill, W. G., & Robertson, A. (1966). The effect of linkage on limits to artificial selection. *Genetics Research*, *8*(3), 269-294. Hölldobler, B., & Wilson, E. O. (1977). The number of queens: An important trait in ant evolution. *Naturwissenschaften*, *64*(1), 8–15.

Hölldobler, B., & Wilson, E. O. (1977). The number of queens: an important trait in ant evolution. *Naturwissenschaften*, *64*(1), 8-15.

Huang, Y. C., Dang, V. D., Chang, N. C., & Wang, J. (2018). Multiple large inversions and breakpoint rewiring of gene expression in the evolution of the fire ant social supergene. *Proceedings of the Royal Society B: Biological Sciences*, *285*, 20180221.

Johnson, E. M. S., Andrews-Larson, C., Keene, K., Melhuish, K., Keller, R., & Fortune, N. (2020). Inquiry and gender inequity in the undergraduate mathematics classroom. Journal for Research in Mathematics Education, 51(4), 504–516.

Jones, B. M., & Robinson, G. E. (2018). Genetic accommodation and the role of ancestral plasticity in the evolution of insect eusociality. *Journal of Experimental Biology*, *221*(23), jeb153163.

Keller, L., & Passera, L. (1989). Size and fat content of gynes in relation to the mode of colony founding in ants (Hymenoptera; Formicidae). *Oecologia*, *80*(2), 236–240.

Keller, L., & Ross, K. G. (1999). Major gene effects on phenotype and fitness: The relative roles of Pgm-3 and Gp-9 in introduced populations of the fire ant solenopsis invicta. *Journal of Evolutionary Biology*, *12*(4), 672–680.

Keller, L., & Ross, K. G. (1993a). Phenotypic basis of reproductive success in a social insect: genetic and social determinants. *Science*, *260*(5111), 1107-1110.

Keller, L., & Ross, K. G. (1993b). Phenotypic plasticity and “cultural transmission” of alternative social organizations in the fire ant Solenopsis invicta. *Behavioral Ecology and Sociobiology*, *33*(2), 121–129.

Keller, Laurent, & Ross, K. G. (1998). Selfish genes: A green beard in the red fire ant. *Nature*, *394*(6693), 573–575.

Laursen, S. L., Hassi, M. L., Kogan, M., & Weston, T. J. (2014). Benefits for women and men of inquiry-based learning in college mathematics: A multi-institution study. *Journal for Research in Mathematics Education*, *45*(4), 406-418.

Leffer, L. (2021, October 8). These mighty shorebirds keep breaking flight records, and you can follow along. *Audubon*. https://www.audubon.org/news/these-mighty-shorebirds-keep-breaking-flight-records-and-you-can-follow-along

Levis, N. A., Serrato-Capuchina, A., & Pfennig, D. W. (2017). Genetic accommodation in the wild: evolution of gene expression plasticity during character displacement. *Journal of Evolutionary Biology*, *30*(9), 1712-1723.

Martinez-Ruiz, C., Pracana, R., Stolle, E., Paris, C. I., Nichols, R. A., & Wurm, Y. (2020). Genomic architecture and evolutionary conflict drive allele-specific expression in the social supergene of the red fire ant. *bioRxiv*, 2020-02.

Moczek, A. P., Sultan, S., Foster, S., Ledón-Rettig, C., Dworkin, I., Nijhout, H. F., ... & Pfennig, D. W. (2011). The role of developmental plasticity in evolutionary innovation. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1719), 2705-2713.

Nehm, R. H., Finch, S. J., & Sbeglia, G. C. (2022). Is active learning enough? The contributions of misconception-focused instruction and active-learning dosage on student learning of evolution. *BioScience*, *72*(11), 1105-1117.

Nijhout, H. F., Kudla, A. M., & Hazelwood, C. C. (2021). Genetic assimilation and accommodation: models and mechanisms. *Current topics in developmental biology*, *141*, 337-369.

Park, S., Jang, J. Y., Chen, Y. C., & Jung, J. (2011). Is pedagogical content knowledge (PCK) necessary for reformed science teaching?: Evidence from an empirical study. *Research in Science Education*, *41*, 245-260.

Pfennig, & Ehrenreich. (2014). Towards a gene regulatory network perspective on phenotypic plasticity, genetic accommodation and genetic assimilation. *Molecular Ecology, 23*(18), 4438-4440. doi:10.1111/mec.12887

Pfennig, D. W., & Murphy, P. J. (2000). Character displacement in polyphenic tadpoles. *Evolution*, *54*(5), 1738-1749.

Piersma, T. (1998). Phenotypic flexibility during migration: optimization of organ size contingent on the risks and rewards of fueling and flight?. *Journal of Avian Biology*, 511-520.

Pigliucci, M., Murren, C. J., & Schlichting, C. D. (2006). Phenotypic plasticity and evolution by genetic assimilation. *Journal of Experimental Biology*, *209*(12), 2362-2367.

Pracana, R., Priyam, A., Levantis, I., Nichols, R. A., & Wurm, Y. (2017). The fire ant social chromosome supergene variant Sb shows low diversity but high divergence from SB. *Molecular Ecology*, *26*(11), 2864–2879.

Ross, K. G. (1997). Multilocus evolution in fire ants: effects of selection, gene flow and recombination. *Genetics*, *145*(4), 961–974.

Ross, K. G., & Keller, L. (1998). Genetic control of social organization in an ant. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(24), 14232–14237.

Ross, K. G., & Shoemaker, D. (2018). Unexpected patterns of segregation distortion at a selfish supergene in the fire ant Solenopsis invicta. *BMC Genetics*, *19*(1), 1–22.

Schussler, E. E., Read, Q., Marbach-Ad, G., Miller, K., & Ferzli, M. (2015). Preparing biology graduate teaching assistants for their roles as instructors: An assessment of institutional approaches. CBE Life Sciences Education, 14(3), 1–11.

Scoville, A. G., & Pfrender, M. E. (2010). Phenotypic plasticity facilitates recurrent rapid adaptation to introduced predators. *Proceedings of the National Academy of Sciences*, *107*(9), 4260-4263.

Schwander, T., Libbrecht, R., & Keller, L. (2014). Supergenes and complex phenotypes. *Current Biology*, *24*(7), R288-R294.

Stains, M., Harshman, J., Barker, M. K., Chasteen, S. V., Cole, R., DeChenne-Peters, S. E., ... & Young, A. M. (2018). Anatomy of STEM teaching in North American universities. *Science*, *359*(6383), 1468-1470.

Stearns, S. C. (1989). Trade-offs in life-history evolution. *Functional ecology*, *3*(3), 259-268.

Stolle, E., Pracana, R., Howard, P., Paris, C. I., Brown, S. J., Castillo-Carrillo, C., Rossiter, S. J., & Wurm, Y. (2019). Degenerative Expansion of a Young Supergene. *Molecular biology and evolution*, *36*(3), 553–561.

Trible, W., & Ross, K. G. (2016). Chemical communication of queen supergene status in an ant. *Journal of evolutionary biology*, *29*(3), 502–513.

Tschinkel, W. R., & Howard, D. F. (1978). Queen replacement in orphaned colonies of the fire ant, Solenopsis invicta. *Behavioral Ecology and Sociobiology*, *3*, 297-310.

Tuttle, E. M., Bergland, A. O., Korody, M. L., Brewer, M. S., Newhouse, D. J., Minx, P., ... & Balakrishnan, C. N. (2016). Divergence and functional degradation of a sex chromosome-like supergene. *Current Biology*, *26*(3), 344-350.

Vargo, E. L., & Fletcher, D. J. C. (1989). On the relationship between queen number and fecundity in polygyne colonies of the fire ant Solenopsis invicta. *Physiological Entomology*, *14*(2), 223–232.

Waddington, C. H. (1953). Genetic assimilation of an acquired character. *Evolution*, 118-126.

Wang, J., Wurm, Y., Nipitwattanaphon, M., Riba-Grognuz, O., Huang, Y. C., Shoemaker, D., & Keller, L. (2013). A Y-like social chromosome causes alternative colony organization in fire ants. *Nature*, *493*(7434), 664–668.

Wellenreuther, M., & Bernatchez, L. (2018). Eco-evolutionary genomics of chromosomal inversions. *Trends in ecology & evolution*, *33*(6), 427-440.

West-Eberhard, M. J. (2003). *Developmental plasticity and evolution* (Vol. 816). Oxford University Press.

Wood, D. P., Holmberg, J. A., Osborne, O. G., Helmstetter, A. J., Dunning, L. T., Ellison, A. R., ... & Papadopulos, A. S. (2023). Genetic assimilation of ancestral plasticity during parallel adaptation to zinc contamination in Silene uniflora. *Nature Ecology &*

Evolution, *7*(3), 414-423.

Yan, Z., Martin, S. H., Gotzek, D., Arsenault, S. V., Duchen, P., Helleu, Q., Riba-Grognuz, O., Hunt, B. G., Salamin, N., Shoemaker, D., Ross, K. G., & Keller, L. (2020). Evolution of a supergene that regulates a trans-species social polymorphism. *Nature Ecology & Evolution*, *4*, 240-249.

CHAPTER TWO

MOLECULAR UNDERPINNINGS OF PLASTICITY AND SUPERGENE-MEDIATED POLYMORPHISM IN FIRE ANT QUEENS[1]

[1] Waugh, A.H., M.A. Catto, S. Kay, S.V. Arsenault, K.G. Ross, and B.G. Hunt. Submitted to *Journal of Evolutionary Biology*, 10/17/24.

Abstract

Characterizing molecular underpinnings of plastic traits and balanced polymorphisms represent two important goals of evolutionary biology. Fire ant gynes (pre-reproductive queens) provide an ideal system to study potential links between these phenomena because they exhibit both supergene-mediated polymorphism and nutritional plasticity in weight and colony-founding behavior. Gynes with the inversion supergene haplotype are lightweight and depend on existing workers to initiate reproduction. Gynes with only the ancestral, non-inverted gene arrangement accumulate more nutrient reserves as adults and, in a distinct colony-founding behavior, initiate reproduction without help from workers. However, when such gynes overwinter in the natal nest they develop an environmentally induced lightweight phenotype and colony-founding behavior, similar to gynes with the inversion haplotype that have not overwintered. To evaluate the extent of shared mechanisms between plasticity and balanced polymorphism in fire ant gyne traits, we assessed whether genes with expression variation linked to overwintering plasticity may be affected by evolutionary divergence between supergene haplotypes. To do so, we first compared transcriptional profiles of brains and ovaries from overwintered and non-overwintered gynes to identify plasticity-associated genes. These genes were enriched for metabolic and behavioral functions. Next, we compared plasticity-associated genes to those differentially expressed by supergene genotype, revealing a significant overlap of the two sets in ovarian tissues. We also identified sequence substitutions between supergene variants of multiple plasticity-associated genes, consistent with a scenario in which an ancestrally plastic phenotype responsive to an environmental condition became increasingly genetically regulated.

Introduction

Phenotypic plasticity refers to the capacity of a single genotype to produce a range of phenotypes in response to different environmental conditions (Stearns, 1989; West-Eberhard, 2003). Plasticity is observed in a wide variety of traits in all domains of life, giving rise to evolutionarily significant morphological, physiological, behavioral, and life history variation (Sommer, 2020; Dupont et al., 2023). When environmental conditions fluctuate, plasticity can permit organisms to respond adaptively to the environment by producing either continuous or discrete trait variation, with the latter resulting in alternative phenotypes. In contrast, alternative phenotypes can also be maintained as part of a genetically regulated balanced polymorphism. Two important aims of evolutionary biology are to unravel the molecular basis of plastic traits and to elucidate the genetic processes that sustain balanced polymorphisms. A promising strategy to achieve both objectives involves studying organisms where similar phenotypes arise independently as either plastic responses to environmental stimuli or as genetically regulated outcomes of balanced polymorphisms.

The red imported fire ant (*Solenopsis invicta*) provides an opportunity to study the interplay of plasticity and genetic regulation in the context of a chromosomal inversion-derived supergene, a taxonomically widespread genomic architecture for the long-term maintenance of alternative morphs in natural populations (Wellenreuther & Bernatchez, 2018; Schwander et al., 2014). In fire ants, polygyny (worker toleration of multiple queens) is a secondary characteristic derived from monogyny (worker toleration of only a single queen; Boomsma et al., 2014). *S. invicta* and multiple congeners exhibit either monogyny or polygyny as determined by a social chromosome bearing an inversion polymorphism that arose relative recently (during the last 0.5 million years) and spread via introgression (Wang et al., 2013; Stolle et al., 2019; Yan et al., 2020; Helleu et al., 2022; Stolle et al., 2022). This inversion polymorphism acts as a supergene system to regulate a suite of morphological and life history traits associated with colony queen number (Wang et al., 2013; Yan et al., 2020; Kay et al., 2022; Chapuisat, 2023).

As part of the regulation of monogyne and polygyne life histories, the fire ant social chromosome regulates the maintenance of two gyne (pre-reproductive queen) ecotypes whose weight accumulation and corresponding colony founding behavior can be predicted by the presence or absence of the derived inversion-carrying supergene allele (*Sb*; Figure 2.1A; Keller & Ross, 1993a, 1993b; DeHeer et al., 1999; Keller & Ross, 1999; DeHeer, 2002; Wang et al., 2013).

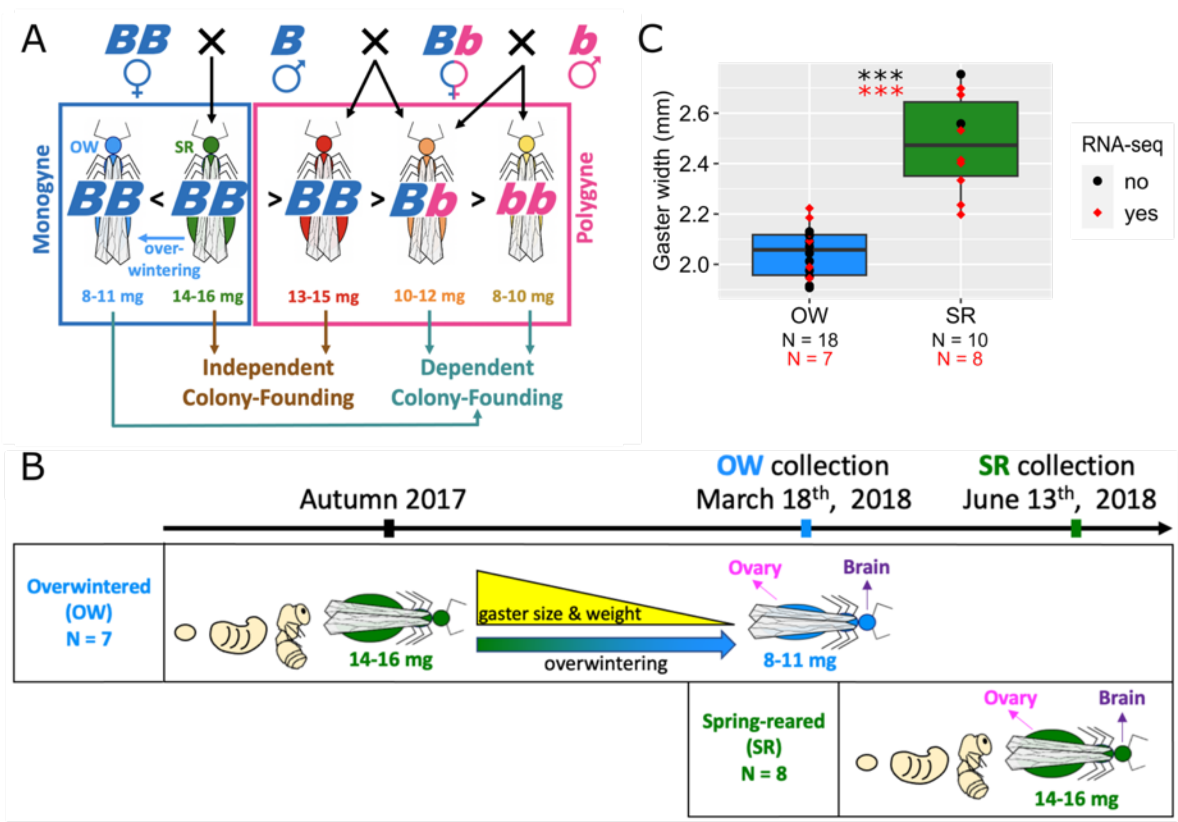


Figure 2.1. Gyne collection timeline, colony founding behavior, and size differences between overwintered (OW) and spring-reared (SR) gynes. (A) The relationship between gyne weight (spring-reared unless otherwise specified), supergene genotype, and colony founding behavior. Parental supergene genotypes of haploid males and diploid mother queens are shown at top. Typical weight ranges for each type of gyne are shown in milligrams with relationships between pairs of gyne types emphasized using greater than and less than symbols (Keller & Ross, 1993b, 1993a; Tschinkel, 1996; DeHeer et al., 1999; Keller & Ross, 1999;

DeHeer, 2002). Supergene genotype is shown in simplified form as *B* = *SB* and *b* = *Sb*. Arrows lead to each gyne type's colony founding mode (independent or dependent). (B) Overwintered gynes (light blue) eclosed in Autumn, 2017 and overwintered in their natal nest before collection in March 2018, when they emerged for their nuptial flight. Spring-reared gynes (green) eclosed in the Spring of 2018 and were collected in June 2018 when they emerged for their nuptial flight. Brain and ovary tissues were extracted from each harvested gyne and used as material for RNA extraction. Sample sizes reflect the final number of biological replicates used for differential gene expression testing following removal of sample OW1 (see methods for details). (C) Boxplots of gyne gaster width for samples measured for this study. Data from gynes used for gene expression analysis are shown with red diamonds as opposed to black circles for gynes measured but not used for gene expression analysis. *S. invicta* gynes homozygous for the ancestral gene arrangement at the supergene (*SB/SB*) embark on mating flights with robust nutrient reserves and proceed to found colonies from scratch in their claustral chamber independently of worker assistance (DeHeer & Tschinkel, 1998; DeHeer et al., 1999; DeHeer, 2002). In contrast, gynes heterozygous for the derived supergene allele (*SB/Sb*) accumulate fewer nutrient reserves before embarking on mating flights and are thus only able to successfully rear brood when initiating egg-laying in the presence of workers (Keller & Ross, 1993a; DeHeer, 2002). Unlike *SB/SB* gynes, *SB/Sb* gynes are accepted by existing polygyne colonies after completing a mating flight. Gynes homozygous for the derived supergene allele (*Sb/Sb*) are also accepted by workers in polygyne colonies but are exceptionally low weight and of low fitness (DeHeer, 2002; Hallar et al., 2007).

The association between gyne weight and mode of colony founding is not unique to fire ants and is observed in many ant species (Keller & Passera, 1989, 1990). Variation in individual gyne weight is substantially influenced by the amount of fat content an individual accrues during development and maturation (Keller & Passera, 1990; Keller & Ross, 1993b, 1993a). Species whose gynes exhibit greater than 40% relative fat content by dry weight typically engage in independent colony founding whereas other species whose gynes accrue less relative fat typically engage in dependent founding (Keller & Passera, 1989). In fire ants, fat content makes up 44% of total dry body weight for independently founding *SB/SB* gynes, roughly 32% for dependently founding *SB/Sb* gynes, and even less for low fitness *Sb/Sb* gynes which seldom succeed in becoming functional queens (Keller & Ross, 1993a; DeHeer et al., 1999; Keller & Ross, 1999; DeHeer, 2002). Thus, fire ant gynes conform to the association between nutrient reserves and mode of colony founding observed across a variety of ant species.

Interestingly, gynes reared in monogyne colonies (all *SB/SB*) that eclose at the end of the mating season typically overwinter in their natal nest and proceed to lose weight over the course of the winter (Figure 2.1B; Fletcher & Blum, 1983; Tschinkel, 1996; DeHeer & Tschinkel, 1998; Helms & Godfrey, 2016). The overwintered *SB/SB* gynes of *S. invicta*, which make up 8-10% of annual gyne biomass produced by a colony (Morrill, 1974), exhibit plasticity in traits that eventually develop to resemble in many ways those of gynes carrying the *Sb* supergene (Tschinkel & Howard, 1978; Fletcher & Blum, 1983; Tschinkel, 1993; Tschinkel, 1996). First, overwintered gynes, like *Sb*-carrying gynes (Keller & Ross, 1993a), exhibit reduced fat content compared to their non-overwintered *SB/SB* counterparts (Tschinkel, 1996; Helms & Godfrey, 2016). Importantly, the fat reserves of overwintered *SB/SB* gynes typically fall below the estimated 40% fat content threshold estimated to be necessary for successful independent founding (Keller & Passera, 1989, Tschinkel, 1996). Moreover, both spring-reared and overwintered gynes exhibit similar head widths in *S. invicta*, suggesting similar overall body sizes at the time of adult eclosion (Helms & Godfrey, 2016). This indicates the reduced weight phenotype of overwintered *SB/SB* gynes occurs as a consequence of phenotypic plasticity in the adult stage and not because overwintered gynes represent a developmental caste polymorphism.

Second, rather than digging a claustral chamber after mating, both overwintered and *Sb*-carrying gynes engage in dependent founding as they attempt to enter and begin egg-laying in an existing nest where workers are present (DeHeer & Tschinkel, 1998; DeHeer, 2002). Overwintered gynes engage in a specific form of dependent colony founding known as queen replacement (Tschinkel & Howard, 1978; Tschinkel, 1996; DeHeer & Tschinkel, 1998). When successful, queen replacement involves a newly mated overwintered fire ant gyne invading a monogyne nest whose queen recently died. In such a case, orphaned workers will accept and tend to the invading overwintered gyne as their new reproductive queen (Tschinkel & Howard, 1978). It was estimated in one study population that 0.7% of colonies per year adopt replacement queens and based on extrapolation, 3% of mature nests may be expected to be headed by overwintered queens (DeHeer & Tschinkel, 1998). Thus, while the individual success rate via queen replacement for individual overwintered gynes is low, at the population level, this reproductive strategy exists as a viable means of yielding reproductive returns from otherwise underweight, low fitness gynes. Like overwintered *SB/SB* gynes, *Sb*-carrying gynes are ill-equipped for independent nest founding (DeHeer, 2002) and invade other nests after embarking on mating flights, in this case being accepted when the colony is either polygyne or queen-less.

Third, overwintered and *Sb*-carrying gynes exhibit maternally incompetent behaviors (Tschinkel, 1996; DeHeer, 2002). When monogyne overwintered gynes (*SB/SB*) have been forced into simulated claustral independent colony founding conditions in the laboratory, they lay eggs, but they do not cluster them as is typical of spring-reared *SB/SB* gynes (Tschinkel, 1996). Overwintered gynes also tend to fail to feed and rear the larvae that hatch from their haphazardly laid eggs (Tschinkel, 1996). Similar maternally incompetent behaviors have been observed among newly mated polygyne *Sb*-carrying queens forced into simulated claustral, independent colony founding conditions in the laboratory (DeHeer, 2002). Maternal incompetence among overwintered *SB/SB* and *Sb*-carrying gynes may occur as a consequence of relaxed selection on brood tending behaviors when queens are in a nutritional state incompatible with independent nest founding.

In this study, we leverage the unique biology of the fire ant system, in which similar alternative phenotypes arise through both plasticity and supergene-mediated balanced polymorphism. We aim to assess the similarity in the underlying mechanisms of plasticity and a genetic polymorphism by investigating whether the genes associated with plastic responses to overwintering also contribute to supergene-mediated variation. Our first goal is to understand the changes in gene expression associated with overwintering by the fire ant queen caste to better understand the gene regulatory mechanisms associated with plasticity in adult weight and colony founding behavior. Toward this end, we compare gene expression profiles of brain and ovary tissues in overwintered queens to those from non-overwintered, spring-reared queens, with both types collected while embarking on mating flights. We selected these two tissues because they play important roles in the behavioral and reproductive physiology linked to queen phenotypes of interest. Our second goal is to determine whether some of the genes involved in mediating environmentally induced plasticity in weight and colony founding behavior may also function in the genetic regulation of these phenotypes.

To provide insight into this prospect, we compare the genes that are differentially expressed by overwintering status to those differentially expressed by supergene genotype in brain and ovary tissues at a comparable life history stage (Arsenault et al., 2020). In order to assess whether genes associated with plasticity in nutrient reserves have been perturbed during the course of supergene evolution, we also compare the genes differentially expressed by overwintering status to those with fixed substitutions between *SB* and *Sb* alleles of the supergene (Martinez-Ruiz et al., 2020). Our results indicate that the supergene perturbs a set of genes with enriched overlap with those differentially expressed in association with nutritional plasticity. Plausibly, via genetic assimilation (West-Eberhard, 2003, 2005), the fire ant supergene may act to maintain the balanced queen weight polymorphism by perturbing the molecular underpinnings of ancestral nutritional plasticity. We discuss evidence consistent with this novel evolutionary role for supergenes and call for future work to formally test the hypothesis of supergene-mediated genetic assimilation in fire ants and other organisms.

Materials and Methods

Sample collection and processing

Solenopsis invicta alate gynes and workers were aspirated from the tops of nests of the monogyne social form and frozen on dry ice in the field on days of mating flights along roadsides in and around Oconee National Forest in Georgia, USA. Spring-reared *SB/SB* gynes were collected on June 13, 2018, in Greene County, Georgia (after any overwintering gynes would have departed on mating flights (Fletcher & Blum, 1983); Figure 2.1A). Monogyne colonies of *S. invicta* in this area do not produce sexuals during the winter months and produce their first and largest pulse of sexuals in early- to mid-spring (Vargo & Fletcher, 1987). Overwintered *SB/SB* gynes were collected on March 18, 2018, in nearby Oglethorpe County, Georgia, prior to the production of non-overwintering adult sexuals in this area (Figure 2.1B). To confirm monogyne social form of colonies, a *Gp-9* PCR assay on pooled individuals was performed (Valles & Porter, 2003). All samples were stored at -80°C until the addition of RNA*later*-ICE, at which point gynes were stored for at least 24 hours at -20°C prior to dissection.

S. invicta gyne and reproductive queen gasters (abdomens) distend as necessary (e.g., to make room for crop, fat body, and ovarian enlargement). Two-dimensional gaster measurements were taken to confirm size differences between randomly sampled overwintered and spring-reared gynes (Spreadsheet 2). Sampled overwintered gyne gasters in our study were significantly smaller than those of spring-reared gynes (One-sided Mann-Whitney U; N = 8 SR, 7 OW; p = 0.0003; Figure 2.1C; Spreadsheet 2), consistent with observations from more extensive sampling efforts of corresponding weight differences (Fletcher & Blum, 1983; Tschinkel, 1996; Helms & Godfrey, 2016), all of which confirm lower nutrient reserves in overwintered gynes.

Dissection and storage of brains and ovaries were performed as described by Arsenault et al. (2020). RNA was extracted from brains using the RNeasy Micro Kit and from ovaries using the RNeasy Mini Kit with DNase treatment (Qiagen, Valencia, CA). Extracted total RNA integrity and concentration were evaluated on an Agilent 2100 bioanalyzer (Supplementary Figure S2.1). Dissected tissue from a single gyne from each colony were used to prepare libraries following the Smart-seq2 protocol developed for low input RNA sequencing applications (Picelli et al., 2014). Based on bioanalyzer readings, approximately 1.6 ng of total RNA was used to make each brain library and approximately 8.0 ng of total RNA was used to make each ovary library. Samples were barcoded and pooled for sequencing at the Georgia Genomics and Bioinformatics Core (Athens, GA) on an Illumina NextSeq sequencer to produce 75 bp, single-end reads. In total, samples from 16 unrelated *SB/SB* gynes, 8 overwintered and 8 spring-reared, from 16 different monogyne nests, yielded 16 brain and 16 ovary RNA-seq libraries for our study.

RNA-seq quality control and alignment

We trimmed reads and performed quality control using *Trim Galore!* v0.6.5. We then used *STAR* v2.7.3a (Dobin et al., 2013) to align reads to the “SINVBB1” genome assembly and associated annotation (GCA_009650705.1; Yan et al., 2020) using the 2-pass alignment procedure. Following alignment, brain libraries contained between 11 and 23 million and ovary libraries between 12 and 27 million uniquely mapped reads. We loaded quantified, gene-level read counts into *edgeR* (Robinson et al., 2009) for subsequent analyses.

For each tissue, we matched the expression cutoff used by Arsenault et al*.,* (2020) by removing genes with fewer than 1.11 counts per million (CPM) in all 15 libraries from subsequent analyses. Of the 16,314 annotated genes in the fire ant genome (GCA_009650705.1), 47% (7,675/16,314) and 49% (8,048/16,314) of genes passed our low count filter in the brain and ovary comparisons respectively. PCA and HCA were performed for each tissue comparison separately using the CPM value for all genes that passed our low count filter. To visualize expression variance and relationships between libraries, HCA was performed with the R package *pheatmap* using the “ward D2” clustering method (Figure S2). Brain and ovary libraries from one individual, OW1, were removed due to outlier behavior in our principal component analysis (PCA) and hierarchical clustering analysis (HCA; Figure S2) that suggested potential low complexity of sequencing libraries.

Differential gene expression

For each tissue, we used *glmQLFTest* in *edgeR* (Robinson et al., 2009) to perform a separate pairwise design to test for differentially expressed genes between sample types. To assess similarity in the transcriptomic effects of overwintering and supergene genotype, we re-analyzed the RNA-seq data from Arsenault et al., (Arsenault et al., 2020) that were also generated from brains and ovaries of *S. invicta* gynes that were aspirated from the top of mounds on days of mating flights in Northeast Georgia, USA; single *SB/SB*, *SB/Sb*, and *Sb/Sb* gynes were collected from eight separate nests to comprise their biological replicates. Within tissue-type pairwise differential expression analyses of these data and the overwintered versus spring-reared data were performed with a false discovery rate (FDR) significance cutoff < 0.05 (Table S8).

Functional enrichment

We performed gene ontology (GO) enrichment analyses using the “elim” method from *topGO* (Alexa & Rahnenfuhrer, 2010). We first used *biomaRt* (Kinsella et al., 2011) to obtain the *Drosophila melanogaster* ortholog from the Ensembl metazoan database. The background set for these tests consisted of all genes passing the low count threshold for each respective tissue. The foreground sets consisted of differentially expressed genes (DEGs) found in each tissue. Not all brain and ovary DEGs had a *Drosophila* ortholog listed in the Ensembl database and thus were not included in these analyses. When *S. invicta* genes had multiple *Drosophila* orthologs, we used the *D. melanogaster* ortholog with the highest sequence identity to the target

S. invicta gene. We present significantly enriched biological process GO terms called at *elimKS* < 0.05 (Spreadsheets S10-S11).

Nucleotide substitutions between supergene alleles

A prior comparison of 20 *SB* and 20 *Sb S. invicta* haploid male genomes from the US and South America identified a set of 96 genes (Spreadsheet S12) for which the position of a fixed nucleotide substitution in the *Sb* allele either affected a regulatory region (3' and/or 5' untranslated region, UTR) or changed the amino acid sequence of the expressed protein (nonsynonymous substitution) (Martinez-Ruiz et al., 2020). We compared this aggregated set of genes with fixed substitutions to the genes differentially expressed by overwintering status to identify genes of interest in the hypothesized genetic assimilation of lightweight and dependently founding gynes via molecular evolution. *InterProScan* v5.61 (Jones et al., 2014) with all available applications was run on protein sequences for these genes of interest. To test whether any such substitutions overlap with functional protein domains, genomic coordinates reported by Martinez-Ruiz et al., (Martinez-Ruiz et al., 2020) were converted to gene coordinates using a custom R script (Spreadsheet S13).

Results

Transcriptomic effects of overwintering

Our analysis revealed major effects of overwintering status on the transcriptomes of *S. invicta* gynes. In principal components analyses (PCA) conducted separately for each tissue type, PC1 explained around 20% of the total transcriptomic variance (Figures 2.2A & 2.2B), with clear separation of samples by overwintering status on PC1. We observed 667 differentially expressed genes (DEGs; FDR < 0.05) by overwintering status in brain tissues (Figure 2.2C) and

1122 DEGs by overwintering status in ovarian tissues (Figure 2.2D). We observed a significant bias toward upregulation among DEGs in the brains of spring-reared relative to overwintered gynes (65%; 431/667; $X^2 = 57.0$; $p = 4.3\times10^{-14}$) but no such bias in ovaries (51%; 573/1122; $X^2 = 0.51$; $p = 0.47$). The overlap among DEGs by overwintering status between tissues was significantly greater than expected by chance, with 144 DEGs common to both tissue types (Fisher's exact test; odds ratio = 1.48; $p = 0.0001$; Figure 2.2E). We found that 83% (120/144; odds ratio = 26.2; $p = 1.3\times10^{-12}$; Figure 2.2F) of overlapping DEGs were consistently up- or down-regulated by overwintering status in each tissue.

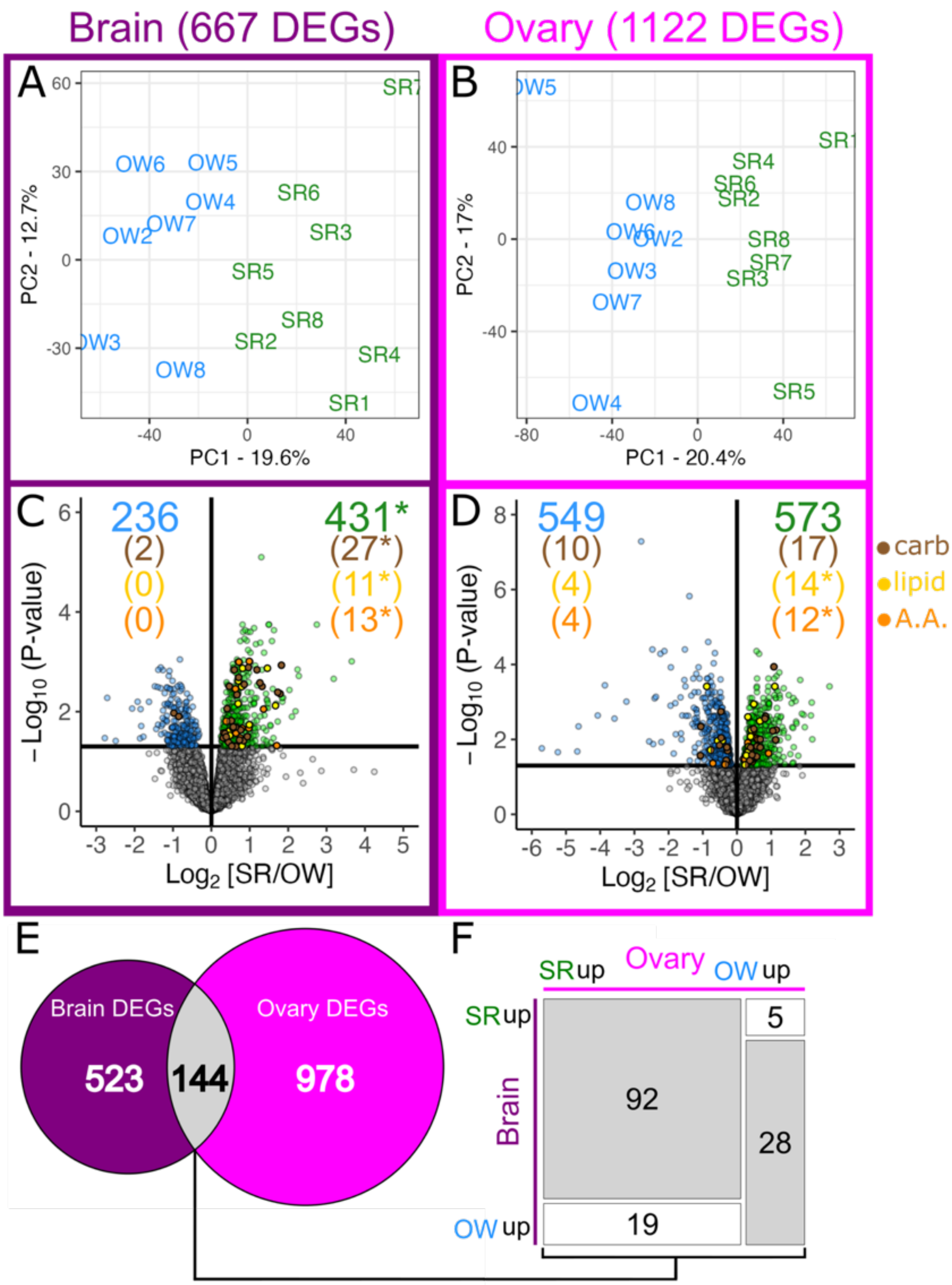


Figure 2.2. Effects of overwintering on gene expression. PCA plots based on normalized gene expression (CPM) for all genes passing the respective expression cutoff for each tissue: brain (A) and ovary (B). Volcano plots of pairwise gene expression differences between overwintered (OW) and spring-reared (SR) fire ant gyne brains (C) and ovaries (D). For each tissue, differentially expressed genes (DEGs; FDR < 0.05) upregulated in SR (green) and OW (blue) gynes are shown. *S. invicta* DEGs whose *D. melanogaster* ortholog (see methods) is annotated to the GO term 'carbohydrate metabolic processes' (GO:0005975, BP) are shown in

brown, ‘cellular lipid catabolic process’ (GO:0044242, BP) in yellow, and ‘alpha-amino acid metabolic process’ (GO:1901605) in orange. The total number of DEGs upregulated in OW and SR gynes are shown in blue and green text respectively. Asterisks (*) on the volcano plot signify significant bias toward down-regulation in the overwintered sample type for each set of genes DEGs annotated to ‘cellular lipid catabolic process’ in the brain (100% (11/11); $X^2 = 11.0$; $p = 0.001$) and ovary (81% (12/14); $X^2 = 5.6$; $p = 0.018$), ‘alpha-amino acid metabolic process’ in the brain (100% (13/13); $X^2 = 13.0$; $p = 0.0003$) and ovary (75% (12/16); $X^2 = 4$; $p = 0.046$), and ‘carbohydrate metabolic process’ in the brain (93% (27/29); $X^2 = 21.6$; $p = 3.4\times10^{-6}$) but not the ovary (63% (17/27); $X^2 = 1.8$; $p = 0.178$). Euler plot Overlap of DEGs found in each tissue (Fisher’s exact test, $p < 0.001$) (E). Mosaic plot showing the directional concordance in DEGs common to the brain and ovary comparisons (F).

To identify candidate functional pathways associated with overwintering status of gynes, we performed Gene Ontology (GO) term enrichment analyses for DEGs (Tables 2.1 & Spreadsheets S2-S7). DEGs by overwintering status in each tissue were enriched for metabolic processes involving three major types of macromolecules, lipids, amino acids, and carbohydrates, with a general pattern of down-regulation in the overwintered sample type (Figure 2.2). DEGs by overwintering status in brains were enriched for several biological processes with relevance to behavior (Tables 2.1 & Spreadsheet S2). DEGs by overwintering status in ovaries were enriched for several biological processes directly related to female reproductive physiology (Tables 2.1 & Spreadsheet S5) and ‘aging’ (Spreadsheet S5). GO enrichment for molecular function (Spreadsheets S3 & S6) and cellular component (Spreadsheets S4 & S7) among brain and ovary overwintering DEGs can be found in the supplementary spreadsheets.

Table 2.1. Top 10 significantly enriched gene ontology biological process terms among DEGs according to overwintering status in the brain (Br) and ovary (Ov).

Tissue	GO term ID	GO Term description	Expressed genes	DEGs	Expected DEGs	*p*-value (elimKS)
Br	GO:0055085	transmembrane transport	345	49	29.51	0.00006
Br	GO:0046835	carbohydrate phosphorylation	9	5	0.77	0.00034
Br	GO:0006006	glucose metabolic process	27	8	2.31	0.00102
Br	GO:0043171	peptide catabolic process	7	4	0.6	0.00126
Br	GO:0019563	glycerol catabolic process	4	3	0.34	0.00203
Br	GO:0006166	purine ribonucleoside salvage	4	3	0.34	0.00203

Br	GO:0000381	regulation of alternative mRNA splicing, via spliceosome	59	12	5.05	0.00243
Br	GO:0006812	monoatomic cation transport	170	25	14.54	0.00254
Br	GO:0051606	detection of stimulus	45	10	3.85	0.00279
Br	GO:0007611	learning or memory	84	15	7.19	0.00288
Ov	GO:0009154	purine ribonucleotide catabolic process	5	4	0.7	0.0015
Ov	GO:0016266	O-glycan processing	5	4	0.7	0.0015
Ov	GO:0035337	fatty-acyl-CoA metabolic process	11	6	1.55	0.0015
Ov	GO:0006633	fatty acid biosynthetic process	37	12	5.2	0.0025
Ov	GO:0045823	positive regulation of heart contraction	3	3	0.42	0.0025
Ov	GO:0050906	detection of stimulus involved in sensory perception	12	6	1.69	0.0027
Ov	GO:0015718	monocarboxylic acid transport	16	7	2.25	0.003
Ov	GO:0030720	oocyte localization involved in germarium-derived egg	9	5	1.27	0.0035
Ov	GO:0044242	cellular lipid catabolic process	62	16	8.72	0.0068
Ov	GO:0009064	glutamine family amino acid metabolic process	27	9	3.8	0.0069

Transcriptomic effects of overwintering relative to supergene genotype

We next compared DEGs by overwintering status in *SB/SB* gynes to DEGs by supergene genotype of spring-reared gynes (eight trios of *SB/SB*, *SB/Sb,* & *Sb/Sb* gynes, each sampled from one of eight polygyne nests; (Arsenault et al., 2020). Of the 1,122 DEGs by overwintering status in ovaries that passed quality control for analysis of genotypic effects, 16% exhibited differential expression between polygyne spring-reared *SB/SB* and *Sb/Sb* gyne ovaries, representing a significantly greater overlap than expected by chance (177/1122; Fisher's exact test; odds ratio = 1.28; p = 0.004; Figure 2.3B). There was no greater overlap than expected by chance between DEGs in brains by overwintering and supergene genotype (Figure 2.3A) or between DEGs in ovaries by overwintering and *SB/SB* versus *SB/Sb* gynes (Figure 2.3B). A full gene expression compendium can be found in Spreadsheet S8.

Among the 177 ovarian DEGs common to overwintering status and *SB/SB* versus *Sb/Sb* supergene homozygotes, 66% (117/177; Fisher's exact test; odds ratio = 3.17; p = 5.5×10^{-4}; Figure 2.3C) showed a discordant pattern of expression. This means they were up-regulated in the lighter-weight sample type in one comparison and down-regulated in the lighter-weight sample type in the other comparison, or *vice versa*. The observation that around two-thirds of the loci influenced by both overwintering and supergene genotype show directional discordance in relation to gyne weight indicates that most of these genes are not solely differentially expressed due to a shared state of nutrient reserve depletion in *Sb/Sb* spring-reared gynes and *SB/SB* overwintered gynes.

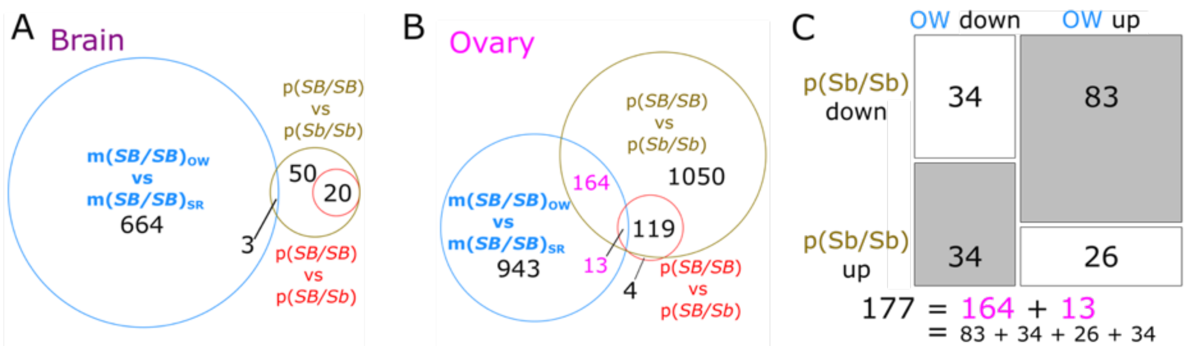


Figure 2.3. DEGs associated with overwintering and supergene genotype. Euler diagram showing overlaps in sets of differentially expressed genes (FDR < 0.05) found in the brain (A) and ovary (B) comparisons of monogyne *SB/SB* overwintered (OW) vs. monogyne *SB/SB* spring-reared (SR) (blue), polygyne *SB/SB* vs. polygyne *SB/Sb* (red), and polygyne *SB/SB* vs polygyne *Sb/Sb* (gold) gynes. (C) Mosaic plot showing enrichment of directionally discordant (gray) and concordant (white) DEGs among the 177 shared ovary DEGs in the overlap of monogyne *SB/SB* OW vs. monogyne *SB/SB* SR and polygyne *SB/SB* vs. polygyne *Sb/Sb* DEGs. Discordant genes were oppositely up- and down-regulated in the smaller-gaster, lightweight sample type (*e.g.,* up in OW & down in *Sb/Sb* gynes) in both comparisons. The designations "m" and "p" denote gyne colony social form of origin (monogyne and polygyne, respectively).

To assess whether the directional discordance in the overlap of ovarian DEGs by gyne weight is part of a larger trend of transcriptome-wide discordance, we tested for a correlation between expression $\log_2$ fold-change values for overwintering and *SB/SB* versus *Sb/Sb* when setting positive values to indicate higher expression level in the heavier gyne type. This revealed significant negative correlations between the expression fold-change contrasts of heavy and light gynes stemming from overwintering status versus alternate supergene allele homozygosity (*SB/SB* versus *Sb/Sb*) in both the brain (n = 7,657; Spearman's rho = -0.32; *p* = 9.7 ´ 10^{-180}; Figure S3A) and ovarian tissues (n = 8,040; Spearman's rho = -0.12; *p* = 6.2 ´ 10^{-27}; Figure S3B). A similar comparison with *SB/SB* versus *SB/Sb* heterozygotes was much weaker but still significant in the brain (n = 7,657; Spearman's rho = -0.04; *p* = 1.4 ´ 10^{-4}; Figure S3C) and not significant in the ovarian tissues (n = 8,040; Spearman's rho = -0.01; *p* = 0.22; Figure S3D). The significant anticorrelations between fold-change contrasts of heavy and light gynes associated with overwintering status and supergene homozygosity (*SB/SB* vs. *Sb/Sb*) suggests the enrichment for discordance we observed in overlapping ovarian DEGs does coincide with global differentiation of the transcriptome with respect to the two different sources of gyne weight variation (environmental and genetic).

Of the 552 genes mapped to the supergene homologous region (*SB*), 273 and 288 had expression data in the brain and ovary in our study, respectively. We found that 3.6% (24/667; Fisher's exact test; odds ratio = 1.01; *p* = 0.91) of DEGs by overwintering status in brains and 3.8% (43/1122; Fisher's Exact; odds ratio = 1.09; *p* = 0.60) of DEGs by overwintering status in ovaries were mapped to the supergene-homologous region. These genes represent candidates for driving the genetic assimilation of reduced gyne weight gain in *Sb*-carrying gynes, as other genomic regions are freely recombining between *SB* and *Sb* genomes (Yan et al., 2020).

Overwintering DEGs with nucleotide substitutions between SB and Sb

To further characterize overwintering DEGs with respect to the fire ant supergene-mediated balanced polymorphism, we integrated data into our study from a prior analysis of fixed differences between *SB* and *Sb* alleles of the supergene (Martinez-Ruiz et al., 2020). We used these data to identify overwintering DEGs in each tissue with nucleotide substitutions that either alter amino acid sequences of encoded proteins or fall within the 3' UTR of transcripts, possibly affecting their expression (Mayr, 2017). In the brain, among the 273 genes in the supergene region with gene expression data in our study, 68 had fixed differences identified between *SB* and *Sb* alleles (Martinez-Ruiz et al., 2020). Six percent (4/68) also exhibited differential expression by overwintering status (Fisher's exact test; odds ratio = 0.65; p = 0.85; Spreadsheet S9). Among these genes, *carbohydrate sulfotransferase 11-like* stood out as an interesting candidate gene underlying plasticity and supergene-mediated polymorphism in fire ant queens because of its pattern of down-regulation in overwintered gynes and its non-synonymous substitution that could yield similar phenotypic effects in *Sb*-carrying gynes. In the ovary, among the 288 genes in the *Sb* supergene region with gene expression data in our study, 69 had fixed differences identified between *SB* and *Sb* alleles. Sixteen percent of these exhibited differential expression by overwintering (11/69, Fisher's exact test; odds ratio = 1.17; p = 0.37; Spreadsheet S9). One of these 11 genes, *SEC23-interacting protein*-like, emerged as an interesting candidate. Its *Drosophila* ortholog, *Phosphatidic Acid Phospholipase A1* (*PAPLA1*), is known to produce fly phenotypes similar to overwintered and *Sb*-carrying queens, including lower egg production, reduced metabolic rates, less fat storage, and decreased glycogen reserves (Galikova et al., 2017; see Discussion). Among the combined 15 overwintering DEGs with fixed differences between their *SB* and *Sb* alleles (Martinez-Ruiz et al., 2020), twelve exhibited these substitutions in 3' UTRs and five (including two with 3'UTR substitutions) exhibited one or more amino acid altering missense substitutions (Tables 2.2 & Spreadsheet S9).

Table 2.2. DEGs by overwintering status with *Sb* substitutions in 3' untranslated regions and/or that affect primary protein sequence.

Gene ID[1]	Gene name	3' UTR subs.[2]	NS subs.[2]	Upreg. OW vs SR[3]	Upreg. *SB/SB* vs *Sb/Sb*	*D. melanogaster* ortholog[4] and FlyBase notes[5]

LOC105203065	*calcium-independent phospholipase A2-gamma-like*	1	2	SR$_{\text{ovary}}$	*SB/SB*$_{\text{ovary}}$	No fly ortholog; *PNPLA8* in mouse; involved in fatty acid hydrolysis
LOC105193134	*carbohydrate sulfotransferase 11-like*	1	1	SR$_{\text{brain}}$	*SB/SB*$_{\text{ovary}}$	*CG13937*; Predicted to be involved in carbohydrate biosynthetic process; expressed in fat body
LOC105194585	*SEC23-interacting protein-like*	0	3	SR$_{\text{ovary}}$	Not significant	*PAPLA1*; enables phospholipase activity; required for the endoplasmic reticulum to Golgi trafficking of a family of G-protein coupled receptors
LOC105194453	*charged multivesicular body protein 7*	0	2	OW$_{\text{ovary}}$	*Sb/Sb*$_{\text{brain}}$	*CG5498*; predicted to be involved in late endosome to vacuole transport
LOC105207412	*peroxisomal membrane protein 11C*	0	1	SR$_{\text{brain}}$	*SB/SB*$_{\text{ovary}}$	*CG33474*; Predicted to be involved in peroxisome fission
LOC105194672	*heat shock 70 kDa protein cognate 5*	1	0	SR$_{\text{ovary}}$	*Sb/Sb*$_{\text{ovary}}$	*Hsc70-5*; Predicted to enable several functions, including ATP hydrolysis activity; mitochondrial protein-transporting ATPase activity; and protein folding chaperone
LOC105199797	*ubiquitin-like domain-containing CTD phosphatase 1*	1	0	OW$_{\text{ovary}}$	*SB/SB*$_{\text{ovary}}$	*Ublcp1*; binds and dephosphorylates the nuclear 26S proteasome; inhibits proteasome activity
LOC105203081	*semaphorin-5B*	1	0	SR$_{\text{ovary}}$	*SB/SB*$_{\text{brain}}$	No fly ortholog; *Sema5b* in mouse; involved in neurogenesis during development
LOC105206526	*dopamine receptor 1*	1	0	OW$_{\text{ovary}}$	*SB/SB*$_{\text{ovary}}$, *SB/SB*$_{\text{brain}}$	*Dop1R1*; Receptor for dopamine; activity mediated by G proteins; involved in memory and learning

[1] Uncharacterized genes and overwintering DEGs with 3' UTR substitutions but no supergene differential expression were excluded (listed in Spreadsheet S9); [2] (Martinez-Ruiz et al., 2020); [3] OW: overwintered *SB/SB,* SR: spring-reared *SB/SB*; [4] Orthologs are from OrthoDB v11 (Kuznetsov et al., 2022); [5] (Thurmond et al., 2019)

Discussion

The purposes of our study were two-fold. First, we aimed to understand the changes in gene expression associated with overwintering in monogyne *S. invicta* to better understand the gene regulatory mechanisms of plasticity in fire ant gyne nutrient reserves and affiliated colony founding behavior. Second, given the resemblance in weight and colony founding behavior between monogyne overwintered gynes (*SB/SB*) and polygyne *Sb*-carrying gynes, we investigated whether the genes associated with overwintering-induced plasticity might also be associated with the genetic regulation of these traits. Although we lack causal evidence linking genotype to phenotype for either overwintering or *Sb*-linked trait variation, several results from our study provide insight into the prospect of genetic assimilation by the fire ant supergene.

Genetic assimilation is the evolutionary process by which a phenotype that initially arises as a response to an environmental condition becomes increasingly genetically regulated over time, such that it is expressed even in the absence of the environmental stimulus (West-Eberhard, 2003, 2005). In other words, through selection on environmentally induced trait variation, a phenotype originally induced by environmental factors can become buffered against environmental variation and thus subject to increased genetic control (Waddington, 1953; West-Eberhard, 2003, 2005; Pigliucci et al., 2006; Moczek et al., 2011; Pfennig & Ehrenreich, 2014; Ehrenreich & Pfennig, 2015; Jones & Robinson, 2018; Nijhout et al., 2021; Wood et al., 2023).

Our finding that ovarian overwintering DEGs overlap more than expected by chance with those observed between *SB/SB* and *Sb/Sb* genotypes appears consistent with a scenario in which the regulatory and structural effects of supergene evolution left the nutritional plasticity of adult gynes intact, perhaps by operating instead on the indirect genetic effects experienced by Sb-carrying gynes as they interact with nurse workers in the polygyne social environment (Arsenault et al., 2023; Majidifar et al., 2024). Contrary to our expectations for this scenario, however, most of the genes differentially expressed by both overwintering status and supergene genotype exhibit directionally discordant expression by gyne weight. Thus, although the supergene and overwintering affect many of the same genes, they affect the expression of these genes differently, consistent with evolutionary changes in the molecular machinery regulating nutrient accumulation and/or food seeking behavior in adult gynes. In either scenario, local adaptation is likely to have played an important role in shaping the genetic variation captured and maintained by the supergene (e.g., Feder et al., 2011).

Three fundamental components of fire ant genetics and life history provide further support to genetic assimilation as the appropriate interpretative framework for our study. First, the reliably lightweight phenotype of *Sb*-carrying gynes stands in contrast to the heavyweight phenotype of spring-reared *SB/SB* gynes produced by both monogyne and polygyne colonies (Keller & Ross, 1993a; DeHeer et al., 1999; Keller & Ross, 1999; DeHeer, 2002) but is consistent with the lightweight phenotype of *SB/SB* gynes after overwintering. This could arise from an increase in the genetic regulation (canalization) of adult gyne nutrient reserve accumulation via effects of *Sb* on nutritional plasticity.

Second, prior support for genetic assimilation to decrease plasticity has come from ancestral trait reconstruction to identify lineages where trait plasticity preceded fixation (e.g., (Heil et al., 2004; Levis & Pfennig, 2016; Jones et al., 2017; Wood et al., 2023). In fire ants, phylogenetic analyses support the monogyne social form as ancestral to the polygyne (Ross & Carpenter, 1991; Boomsma et al., 2014). Thus, it is most parsimonious to assume plasticity in gyne weight evident among monogyne overwintered gynes is ancestral to the emergence of the relatively young supergene in this species (Helleu et al., 2022).

Third, chromosome structural rearrangements, such as chromosomal inversions, commonly underpin ecologically relevant polymorphisms in complex traits (Wellenreuther & Bernatchez, 2018; Harringmeyer & Hoekstra, 2022; Chapuisat, 2023) and represent a genomic architecture primed to facilitate genetic assimilation. Genetic assimilation can occur via regulatory and coding sequence evolution since either has the potential to alter or disrupt molecular machinery underlying plasticity to buffer against environmental variation (Scoville & Pfrender, 2010; Ehrenreich & Pfennig, 2015; Levis et al., 2017; Wood et al., 2023). Chromosomal inversions suppress meiotic crossover events and thus can promote the accumulation and fixation of mutations that affect copy numbers and sequences of non-coding, regulatory, and protein coding elements (Hill & Robertson, 1966; Feder et al., 2011; Bachtrog, 2013; Pracana et al., 2017; Wellenreuther & Bernatchez, 2018; Faria et al., 2019; Stolle et al., 2019; Fontana et al., 2020; Martinez-Ruiz et al., 2020). Meiotic crossover suppression also acts to maintain allelic combinations in tight linkage disequilibrium (Wellenreuther & Bernatchez, 2018). Since the loss of plasticity during genetic assimilation can be deleterious if environmental conditions continue to fluctuate, maintenance of high linkage disequilibrium among alleles involved in genetic assimilation can provide a means by which genetically assimilated phenotypes persist in a population amidst environmental fluctuation. Thus, immediate and long-term effects of inversion polymorphisms offer opportunities for selection to act on the genetic machinery underpinning phenotypic plasticity in favor of the stable production of alternative phenotypes and life histories under increased genetic regulation.

We identified some genes that are differentially expressed between spring-reared and overwintered gynes that exhibit fixed differences between *SB* and *Sb* supergene alleles (Martinez-Ruiz et al., 2020). One manner in which genetic assimilation could produce a pattern of discordant expression with respect to gyne weight is through substitutions in 3’ UTRs that influence gene expression levels by altering *cis*-regulatory element sequences or mRNA stability (Mayr, 2017). We identified six genes in the supergene region with 3’ UTR substitutions and differential expression by both supergene genotype and overwintering status, three of which exhibit directional discordance in expression by gyne weight. These genes could contribute to genetically assimilated phenotypes of *Sb*-carriers if the substitution in the 3' UTR impacts the activity of transcriptional regulators, disrupts machinery underlying phenotypic plasticity, and results in a phenotype less responsive to variable environmental stimuli.

One gene showing directional discordance by gyne weight in ovarian tissues and a 3'-UTR substitution between *SB* and *Sb* is the G protein coupled receptor (GPCR) *Dopamine receptor 1* (Dop1R1), which exhibited reduced expression in lightweight *SB/SB* overwintered gynes and elevated expression in spring-reared *Sb*-carrying gynes in their respective comparisons to heavyweight *SB/SB* spring-reared gynes. Many GPCRs, including several dopamine receptors, are differentially expressed in response to starvation in *Drosophila* (Ko et al., 2015). Dopaminergic neurons in mushroom bodies of the brain have also been shown to mediate food seeking behavior in *Drosophila* (Landayan et al., 2018; Tsao et al., 2018) and Dop1R1 in particular has been shown to mediate ethanol and methamphetamine intake preference of flies (Kanno et al., 2021). Fire ant queens actively seek out fecundity amplifying excretions from late stage larvae (Cassill & Vinson, 2007), which makes this an interesting candidate pathway by which food seeking behavior could come to differ by supergene genotype in fire ants. If the 3' UTR substitution in *Sb* of this gene causes the observed relative increase in its expression, it could contribute causally to genetic assimilation

Genes displaying concordant expression levels by gyne weight according to overwintering and supergene genotype could also contribute to genetic assimilation but via a different mechanism. These genes could contribute to assimilated phenotypes of *Sb*-carriers if a nonsynonymous substitution directly impacts an encoded protein's structure and function. Like the mechanism proposed for discordant DEGs with 3' UTRs, changes to protein structure and function could also disrupt physiological machinery for plasticity. One gene that fits this pattern, *calcium-independent phospholipase A2-gamma-like,* may play a direct role in the metabolism of fire ants based on its phospholipid metabolic activity in mammals (Kita et al., 2019), and this gene exhibits a nonsynonymous substitution in an annotated functional domain. A second gene with concordant expression by gyne weight and a nonsynonymous substitution in an annotated functional domain is *carbohydrate sulfotransferase 11-like*, which is predicted to be involved in carbohydrate synthetic processes.

In our study, the gene *SEC23-interacting protein-like* is particularly notable for exhibiting differential expression by overwintering status and having three nonsynonymous substitutions in its protein coding region (the greatest number we observed for an overwintering DEG). This gene is the ortholog of the gene encoding *Phosphatidic Acid Phospholipase A1* (*PAPLA1*) in *Drosophila melanogaster* (Gáliková et al., 2017). Two of the nonsynonymous substitutions in this gene occur within an intrinsically disordered region of the protein, which may be involved in its role in cell signaling (Wright & Dyson, 2015). The pattern of down-regulation in lightweight overwintered gynes and amino-acid changing substitutions in *Sb*-carrying gynes that we observe logically positions this gene as a candidate for genetic assimilation via molecular evolution. Remarkably, PAPLA1 deficiency through genetic perturbation in flies has been shown to cause reduced rates of egg production, lower metabolic rates, reduced fat storage, and reduced glycogen reserves (Gáliková et al., 2017). All of these phenotypic states bear striking similarity to those of *Sb*-carrying *S. invicta* gynes (DeHeer, 2002). Further, the reduced fecundity of *PAPLA1* mutants (Gáliková et al., 2017) is driven in part by egg chamber degeneration, a phenotype that also occurs in response to nutritional shortage (McCall, 2004), thus demonstrating the gene’s direct potential to canalize a plasticity-induced phenotype through mutation. Intriguingly, *PAPLA1* fly mutants develop into normally sized adults but have both lower food intake and energy expenditure than normal flies (Gáliková et al., 2017). Although the effects of the supergene substitutions in the primary protein sequence of this gene in *S. invicta* remain unknown, such a gene would seem to be an ideal candidate for optimizing colony energy investment in adult fire ant gynes.

Several limitations of our study design should be considered when interpreting our results. First, sampled overwintered gynes are older than their non-overwintered counterparts. Thus, overwintering DEGs are likely to be shaped in part by differences in the ages of *SB/SB* spring-reared and overwintered gynes (Lucas et al., 2017), consistent with enrichment of the GO term ‘aging’ we observed. Second, overwintered gynes may have experienced greater environmental stress than spring-reared gynes, which could have long-term consequences on gene expression. Third, *SB/SB* overwintered gynes have accumulated and then depleted nutrient reserves while *Sb*-carrying spring-reared gynes have accumulated fewer nutrient reserves, which may result in physiological differences that directly influence variation in gene activity. Finally, it is possible our samples harbor genetic variation associated with variation in nutrient metabolism. Though we cannot completely rule out the effects of such variation, we did attempt to capture genetic diversity in a balanced manner by sampling only one gyne of each sample type from a given nest. In the future, these limitations could be overcome by testing for associations between genetic variants of *SB* and rates of individual nutrient accumulation in a laboratory experiment and using this information as a point of comparison for genetic differentiation of *SB* and *Sb*. This would provide further insight into the prospect for assimilation by the fire ant supergene and the mechanisms underlying assimilation.

Chromosome structural rearrangements have garnered recent interest as a widespread genomic architecture facilitating the evolution of complex multigenic trait polymorphisms (Wellenreuther & Bernatchez, 2018; Rubenstein et al., 2019; Harringmeyer & Hoekstra, 2022; Kay et al., 2022). However, genetic assimilation as a potential means by which supergenes come to regulate production of a discrete supergene-carrying alternative morph has received little attention. The initial chromosome structural rearrangement and subsequent sequence divergence in its non-recombining region create opportunities for mechanisms underlying plasticity in a trait to be disrupted, resulting in a less plastic phenotype. Furthermore, linkage disequilibrium between supergene loci creates an opportunity to maintain assimilated phenotypes as part of a suite of traits that make up complex polymorphisms within species. Our study highlights the utility of profiling genes with plasticity-associated expression to gain insight into potential assimilation by inversion polymorphisms. Future research should formally test the hypothesis of supergene-mediated genetic assimilation in fire ants and other organisms.

References

Alexa, & Rahnenfuhrer. (2010). topGO: Enrichment Analysis for Gene Ontology: Bioconductor package.

Arsenault, King, Kay, Lacy, Ross, & Hunt. (2020). Simple inheritance, complex regulation: Supergene-mediated fire ant queen polymorphism. *Molecular Ecology, 29*(19), 3622-3636. doi:10.1111/mec.15581

Arsenault, Riba-Grognuz, Shoemaker, Hunt, & Keller. (2023). Direct and indirect genetic effects of a social supergene. *Molecular Ecology, 32*(5), 1087-1097. doi:https://doi.org/10.1111/mec.16830

Bachtrog. (2013). Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nature Reviews Genetics, 14*(2), 113-124. doi:10.1038/nrg3366

Boomsma, Huszár, & Pedersen. (2014). The evolution of multiqueen breeding in eusocial lineages with permanent physically differentiated castes. *Animal Behaviour, 92*, 241-252. doi:https://doi.org/10.1016/j.anbehav.2014.03.005

Cassill, & Vinson. (2007). Effects of Larval Secretions on Queen Fecundity in the Fire Ant. *Annals of the Entomological Society of America, 100*(2), 327-332. doi:10.1603/0013-8746(2007)100[327:Eolsoq]2.0.Co;2

Chapuisat. (2023). Supergenes as drivers of ant evolution. *Myrmecological News, 33*, 1-18. doi:10.25849/myrmecol.news_033:001

DeHeer. (2002). A comparison of the colony-founding potential of queens from single- and multiple-queen colonies of the fire ant *Solenopsis invicta*. *Animal Behaviour, 64*(4), 655-661. doi:10.1006/anbe.2002.3095

DeHeer, Goodisman, & Ross. (1999). Queen dispersal strategies in the multiple-queen form of the fire ant *Solenopsis invicta*. *The American Naturalist, 153*(6), 660-675. doi:10.1086/303205

DeHeer, & Tschinkel. (1998). The success of alternative reproductive tactics in monogyne populations of the ant *Solenopsis invicta*: Significance for transitions in social organization. *Behavioral Ecology, 9*(2), 130-135. doi:10.1093/beheco/9.2.130

Dobin, Davis, Schlesinger, Drenkow, Zaleski, Jha, . . . Gingeras. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics, 29*(1), 15-21. doi:10.1093/bioinformatics/bts635

Dupont, Thierry, Zinger, Legrand, & Jacob. (2023). Beyond reaction norms: the temporal dynamics of phenotypic plasticity. *Trends in Ecology & Evolution*. doi:10.1016/j.tree.2023.08.014

Ehrenreich, & Pfennig. (2015). Genetic assimilation: a review of its potential proximate causes and evolutionary consequences. *Annals of Botany, 117*(5), 769-779. doi:10.1093/aob/mcv130

Faria, Johannesson, Butlin, & Westram. (2019). Evolving Inversions. *Trends in Ecology & Evolution, 34*(3), 239-248. doi:10.1016/j.tree.2018.12.005

Feder, Gejji, Powell, & Nosil. (2011). ADAPTIVE CHROMOSOMAL DIVERGENCE DRIVEN BY MIXED GEOGRAPHIC MODE OF EVOLUTION. *Evolution, 65*(8), 2157-2170. doi:10.1111/j.1558-5646.2011.01321.x

Fletcher, & Blum. (1983). The inhibitory pheromone of queen fire ants: effects of disinhibition on dealation and oviposition by virgin queens. *Journal of Comparative Physiology ? A, 153*(4), 467-475. doi:10.1007/BF00612601

Fontana, Chang, Chang, Lee, Dang, & Wang. (2020). The fire ant social supergene is characterized by extensive gene and transposable element copy number variation. *Molecular Ecology, 29*(1), 105-120. doi:https://doi.org/10.1111/mec.15308

Gáliková, Klepsatel, Münch, & Kühnlein. (2017). Spastic paraplegia-linked phospholipase PAPLA1 is necessary for development, reproduction, and energy metabolism in Drosophila. *Scientific Reports, 7*(1), 46516. doi:10.1038/srep46516

Hallar, Krieger, & Ross. (2007). Potential cause of lethality of an allele implicated in social evolution in fire ants. *Genetica, 131*(1), 69-79. doi:10.1007/s10709-006-9114-5

Harringmeyer, & Hoekstra. (2022). Chromosomal inversion polymorphisms shape the genomic landscape of deer mice. *Nature Ecology & Evolution*. doi:10.1038/s41559-022-01890-0

Heil, Greiner, Meimberg, Krüger, Noyer, Heubl, . . . Boland. (2004). Evolutionary change from induced to constitutive expression of an indirect plant resistance. *Nature, 430*(6996), 205-208. doi:10.1038/nature02703

Helleu, Roux, Ross, & Keller. (2022). Radiation and hybridization underpin the spread of the fire ant social supergene. *Proc Natl Acad Sci U S A, 119*(34), e2201040119. doi:10.1073/pnas.2201040119

Helms, & Godfrey. (2016). Dispersal Polymorphisms in Invasive Fire Ants. *Plos One, 11*(4), e0153955. doi:10.1371/journal.pone.0153955

Hill, & Robertson. (1966). The effect of linkage on limits to artificial selection. *Genetical Research, 8*(3), 269-294. doi:10.1017/S0016672300010156

Jones, Binns, Chang, Fraser, Li, McAnulla, . . . Hunter. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics, 30*(9), 1236-1240. doi:10.1093/bioinformatics/btu031

Jones, Kingwell, Wcislo, & Robinson. (2017). Caste-biased gene expression in a facultatively eusocial bee suggests a role for genetic accommodation in the evolution of eusociality. *Proceedings of the Royal Society B-Biological Sciences, 284*(1846), 20162228. doi:10.1098/rspb.2016.2228

Jones, & Robinson. (2018). Genetic accommodation and the role of ancestral plasticity in the evolution of insect eusociality. *Journal of Experimental Biology, 221*(23), jeb153163. doi:10.1242/jeb.153163

Kanno, Hiramatsu, Kondo, Tanimoto, & Ichinose. (2021). Voluntary intake of psychoactive substances is regulated by the dopamine receptor Dop1R1 in Drosophila. *Scientific Reports, 11*(1). doi:10.1038/s41598-021-82813-0

Kay, Helleu, & Keller. (2022). Iterative evolution of supergene-based social polymorphism in ants. *Philosophical Transactions of the Royal Society B: Biological Sciences, 377*(1856), 20210196. doi:doi:10.1098/rstb.2021.0196

Keller, & Passera. (1989). Size and fat content of gynes in relation to the mode of colony founding in ants (Hymenoptera; Formicidae). *Oecologia, 80*(2), 236-240. doi:10.1007/BF00380157

Keller, & Passera. (1990). Fecundity of ant queens in relation to their age and the mode of colony founding. *Insectes Sociaux, 37*(2), 116-130.

Keller, & Ross. (1993a). Phenotypic basis of reproductive success in a social insect: Genetic and social determinants. *Science, 260*(5111), 1107-1110. doi:10.1126/science.260.5111.1107

Keller, & Ross. (1993b). Phenotypic plasticity and “cultural transmission” of alternative social organizations in the fire ant *Solenopsis invicta*. *Behavioral Ecology and Sociobiology, 33*(2), 121-129. doi:10.1007/BF00171663

Keller, & Ross. (1999). Major gene effects on phenotype and fitness: The relative roles of Pgm-3 and Gp-9 in introduced populations of the fire ant *Solenopsis invicta*. *Journal of Evolutionary Biology, 12*(4), 672-680.

Kinsella, Kahari, Haider, Zamora, Proctor, Spudich, . . . Flicek. (2011). Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database, 2011*(0), bar030-bar030. doi:10.1093/database/bar030

Kita, Shindou, & Shimizu. (2019). Cytosolic phospholipase A(2) and lysophospholipid acyltransferases. *Biochim Biophys Acta Mol Cell Biol Lipids, 1864*(6), 838-845. doi:10.1016/j.bbalip.2018.08.006

Ko, Root, Lindsay, Zaninovich, Shepherd, Wasserman, . . . Wang. (2015). Starvation promotes concerted modulation of appetitive olfactory behavior via parallel neuromodulatory circuits. *eLife, 4*, e08298. doi:10.7554/eLife.08298

Kuznetsov, Tegenfeldt, Manni, Seppey, Berkeley, Kriventseva, & Zdobnov. (2022). OrthoDB v11: annotation of orthologs in the widest sampling of organismal diversity. *Nucleic Acids Research, 51*(D1), D445-D451. doi:10.1093/nar/gkac998

Landayan, Feldman, & Wolf. (2018). Satiation state-dependent dopaminergic control of foraging in Drosophila. *Scientific Reports, 8*(1). doi:10.1038/s41598-018-24217-1

Levis, & Pfennig. (2016). Evaluating ‘Plasticity-First’ Evolution in Nature: Key Criteria and Empirical Approaches. *Trends in Ecology & Evolution, 31*(7), 563-574. doi:10.1016/j.tree.2016.03.012

Levis, Serrato-Capuchina, & Pfennig. (2017). Genetic accommodation in the wild: evolution of gene expression plasticity during character displacement. *Journal of Evolutionary Biology, 30*(9), 1712-1723. doi:10.1111/jeb.13133

Lucas, Romiguier, & Keller. (2017). Gene expression is more strongly influenced by age than caste in the ant Lasius niger. *Molecular Ecology, 26*(19), 5058-5073. doi:https://doi.org/10.1111/mec.14256

Majidifar, V., Psalti, M. N., Coulm, M., Fetzer, E., Teggers, E. M., Rotering, F., Grünewald, J., Mannella, L., Reuter, M., Unte, D. & Libbrecht, R. 2024. Ontogeny of superorganisms: Social control of queen specialization in ants. *Functional Ecology* **38**: 1044-1060.

Martinez-Ruiz, Pracana, Stolle, Paris, Nichols, & Wurm. (2020). Genomic architecture and evolutionary antagonism drive allelic expression bias in the social supergene of red fire ants. *eLife, 9*, e55862. doi:10.7554/eLife.55862

Mayr. (2017). Regulation by 3′-Untranslated Regions. *Annual Review of Genetics, 51*(1), 171-194. doi:10.1146/annurev-genet-120116-024704

McCall. (2004). Eggs over easy: cell death in the Drosophila ovary. *Developmental Biology, 274*(1), 3-14. doi:https://doi.org/10.1016/j.ydbio.2004.07.017

Moczek, Sultan, Foster, Ledón-Rettig, Dworkin, Nijhout, . . . Pfennig. (2011). The role of developmental plasticity in evolutionary innovation. *Proceedings. Biological sciences / The Royal Society, 278*(1719), 2705-2713. doi:10.1098/rspb.2011.0971

Morrill. (1974). Production and Flight of Alate Red Imported Fire Ants. *Environmental Entomology, 3*(2), 265-271. doi:10.1093/ee/3.2.265

Nijhout, Kudla, & Hazelwood. (2021). Genetic assimilation and accommodation: Models and mechanisms. In S. F. Gilbert (Ed.), *Current Topics in Developmental Biology* (Vol. 141, pp. 337-369): Academic Press.

Pfennig, & Ehrenreich. (2014). Towards a gene regulatory network perspective on phenotypic plasticity, genetic accommodation and genetic assimilation. *Molecular Ecology, 23*(18), 4438-4440. doi:10.1111/mec.12887

Picelli, Faridani, Björklund, Winberg, Sagasser, & Sandberg. (2014). Full-length RNA-seq from single cells using Smart-seq2. *Nature protocols, 9*(1), 171-181. doi:10.1038/nprot.2014.006

Pigliucci, Murren, & Schlichting. (2006). Phenotypic plasticity and evolution by genetic assimilation. *Journal of Experimental Biology, 209*(Pt 12), 2362-2367. doi:10.1242/jeb.02070

Pracana, Levantis, Martínez-Ruiz, Stolle, Priyam, & Wurm. (2017). Fire ant social chromosomes: Differences in number, sequence and expression of odorant binding proteins. *Evolution Letters, 1*(4), 199-210. doi:10.1002/evl3.22

Robinson, McCarthy, & Smyth. (2009). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics, 26*(1), 139-140. doi:10.1093/bioinformatics/btp616

Ross, & Carpenter. (1991). Phylogenetic analysis and the evolution of queen number in eusocial Hymenoptera. *Journal of Evolutionary Biology, 4*(1), 117-130.

Rubenstein, Ågren, Carbone, Elde, Hoekstra, Kapheim, . . . Hofmann. (2019). Coevolution of Genome Architecture and Social Behavior. *Trends in Ecology & Evolution, 34*(9), 844-855. doi:10.1016/j.tree.2019.04.011

Scoville, & Pfrender. (2010). Phenotypic plasticity facilitates recurrent rapid adaptation to introduced predators. *Proceedings of the National Academy of Sciences, 107*(9), 4260-4263. doi:10.1073/pnas.0912748107

Sommer. (2020). Phenotypic Plasticity: From Theory and Genetics to Current and Future Challenges. *Genetics, 215*(1), 1-13. doi:10.1534/genetics.120.303163

Stearns. (1989). The Evolutionary Significance of Phenotypic Plasticity. *BioScience, 39*(7), 436-445. doi:10.2307/1311135

Stolle, Pracana, Howard, Paris, Brown, Castillo-Carrillo, . . . Wurm. (2019). Degenerative expansion of a young supergene. *Mol Biol Evol, 36*(3), 553-561. doi:10.1093/molbev/msy236

Stolle, Pracana, López-Osorio, Priebe, Hernández, Castillo-Carrillo, . . . Wurm. (2022). Recurring adaptive introgression of a supergene variant that determines social organization. *Nature Communications, 13*(1), 1180. doi:10.1038/s41467-022-28806-7

Thurmond, Goodman, Strelets, Attrill, Gramates, Marygold, . . . the FlyBase. (2019). FlyBase 2.0: the next generation. *Nucleic Acids Research, 47*(D1), D759-D765. doi:10.1093/nar/gky1003

Tsao, Chen, Lin, Yang, & Lin. (2018). Drosophila mushroom bodies integrate hunger and satiety signals to control innate food-seeking behavior. *eLife, 7*. doi:10.7554/elife.35264

Tschinkel. (1993). Resource allocation, brood production and cannibalism during colony founding in the fire ant, Solenopsis invicta. *Behavioral Ecology and Sociobiology, 33*(4), 209-223. doi:10.1007/bf02027118

Tschinkel. (1996). A newly-discovered mode of colony founding among fire ants. *Insectes Sociaux, 43*(3), 267-276. doi:10.1007/BF01242928

Tschinkel, & Howard. (1978). Queen replacement in orphaned colonies of the fire ant, Solenopsis invicta. *Behavioral Ecology and Sociobiology, 3*(3), 297-310. doi:10.1007/bf00296315

Valles, & Porter. (2003). Identification of polygyne and monogyne fire ant colonies (Solenopsis invicta) by multiplex PCR of Gp-9 alleles. *Insectes Sociaux, 50*(2), 199-200. doi:10.1007/s00040-003-0662-8

Vargo, & Fletcher. (1987). Effect of queen number on the production of sexuals in natural populations of the fire ant, Solenopsis invicta. *Physiological Entomology, 12*(1), 109-116. doi:10.1111/j.1365-3032.1987.tb00729.x

Waddington. (1953). Genetic Assimilation of an Acquired Character. *Evolution, 7*(2), 118-126. doi:10.1111/j.1558-5646.1953.tb00070.x

Wang, Wurm, Nipitwattanaphon, Riba-Grognuz, Huang, Shoemaker, & Keller. (2013). A Y-like social chromosome causes alternative colony organization in fire ants. *Nature, 493*(7434), 664-668. doi:10.1038/nature11832

Wellenreuther, & Bernatchez. (2018). Eco-evolutionary genomics of chromosomal inversions. *Trends in Ecology & Evolution, 33*(6), 427-440. doi:10.1016/j.tree.2018.04.002

West-Eberhard. (2003). *Developmental Plasticity and Evolution*. Oxford: Oxford University Press.

West-Eberhard, M. J. (2005). Developmental plasticity and the origin of species differences. *Proceedings of the National Academy of Sciences, 102*(Suppl. 1), 6543-6549. doi:10.1073/pnas.0501844102

Wood, Holmberg, Osborne, Helmstetter, Dunning, Ellison, . . . Papadopulos. (2023). Genetic assimilation of ancestral plasticity during parallel adaptation to zinc contamination in Silene uniflora. *Nature Ecology & Evolution, 7*(3), 414-423. doi:10.1038/s41559-022-01975-w

Wright, & Dyson. (2015). Intrinsically disordered proteins in cellular signalling and regulation. *Nature Reviews Molecular Cell Biology, 16*(1), 18-29. doi:10.1038/nrm3920

Yan, Martin, Gotzek, Arsenault, Duchen, Helleu, . . . Keller. (2020). Evolution of a supergene that regulates a trans-species social polymorphism. *Nat Ecol Evol, 4*(2), 240-249. doi:10.1038/s41559-019-1081-1

Supplementary Materials

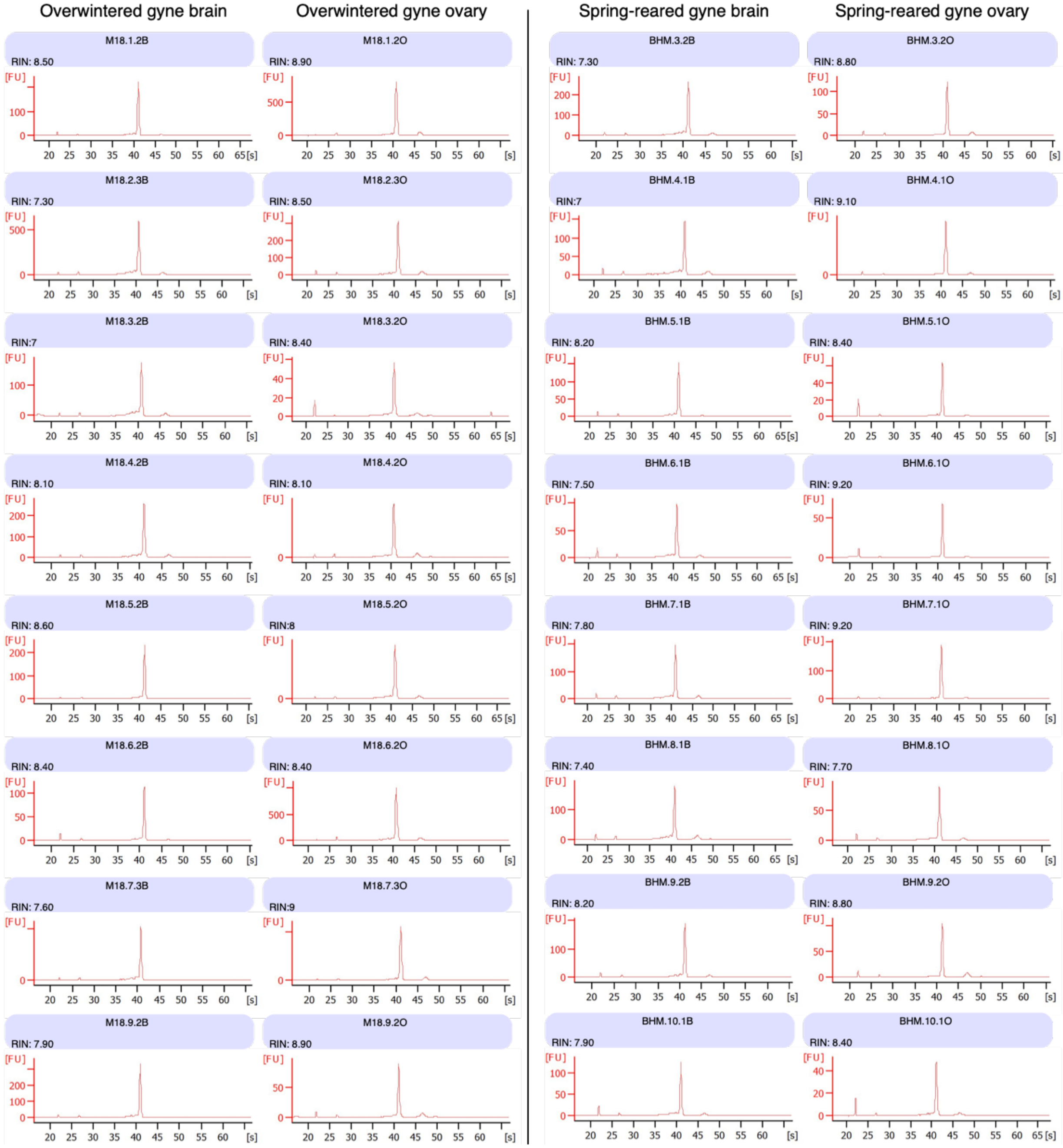


Figure S2.1. Bioanlyzer 2100 trace files for quality control of overwintered and spring reared RNA-seq libraries. Each trace represents a single sample with fluorescent units (FU) on the y axes and time measured in seconds (s) on the x axes. RNA Integrity Number (RIN) (scale: 1-10) is also shown. Higher RIN values indicate more intact RNA. Corresponding sample IDs for each trace file ID shown in purple boxes can be found in ST7.

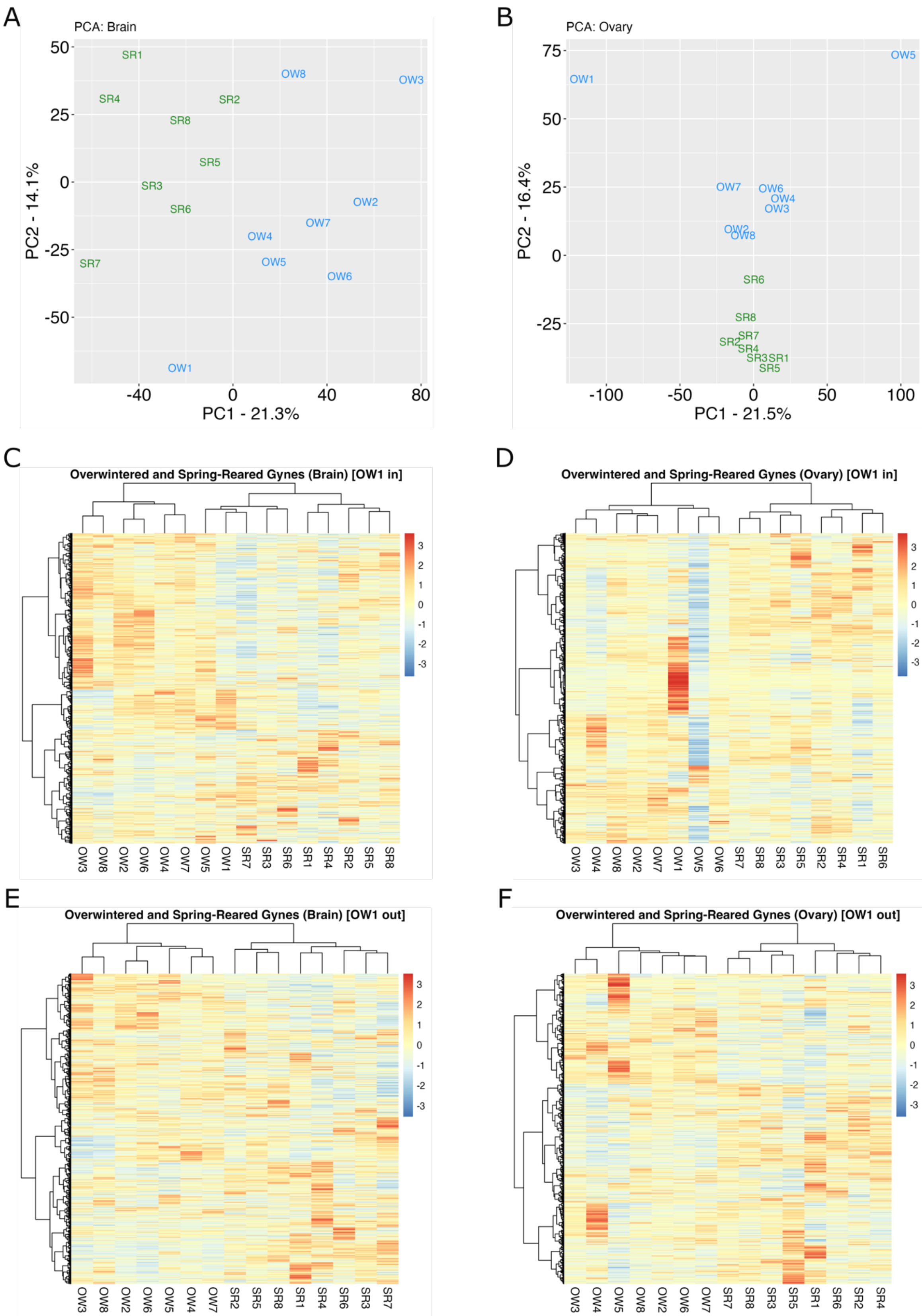

A
PCA: Brain
PC2 - 14.1%
PC1 - 21.3%
SR1
SR4
SR2
SR8
SR5
SR3
SR6
SR7
OW8
OW3
OW2
OW7
OW4
OW5
OW6
OW1
B
PCA: Ovary
PC2 - 16.4%
PC1 - 21.5%
OW5
OW1
OW7
OW6
OW4
OW3
OW2
OW8
SR6
SR8
SR7
SR2
SR4
SR3
SR1
SR5
C
Overwintered and Spring-Reared Gynes (Brain) [OW1 in]
OW3 OW8 OW2 OW6 OW4 OW7 OW5 OW1 SR7 SR3 SR6 SR1 SR4 SR2 SR5 SR8
D
Overwintered and Spring-Reared Gynes (Ovary) [OW1 in]
OW3 OW4 OW8 OW2 OW7 OW1 OW5 OW6 SR7 SR8 SR3 SR5 SR2 SR4 SR1 SR6
E
Overwintered and Spring-Reared Gynes (Brain) [OW1 out]
OW3 OW8 OW2 OW6 OW5 OW4 OW7 SR2 SR5 SR8 SR1 SR4 SR6 SR3 SR7
F
Overwintered and Spring-Reared Gynes (Ovary) [OW1 out]
OW3 OW4 OW5 OW8 OW2 OW6 OW7 SR7 SR8 SR3 SR5 SR1 SR6 SR2 SR4

Figure S2.2. Principal components and hierarchical clustering analyses with outlier OW1 retained. PCA plots including individual OW1 based on normalized gene expression (CPM) for all genes passing the respective expression cutoff for each tissue: brain (A) and ovary (B). Hierarchical clustering analyses (HCA) and heatmaps including data from OW1 based on normalized gene expression (CPM) for all genes passing the respective expression cutoff for each tissue: brain (C) and ovary (D). All data from individual OW1 were removed from all other analyses in this paper. HCA and heatmaps excluding data from OW1 based on normalized gene expression (CPM) for all genes passing the respective expression cutoff for each tissue: brain (E) and ovary (F). All HCA used Ward D2 clustering method. For each heatmap, colors represent Z-scores generated from CPM values indicating up- and down-regulation of genes in each row.

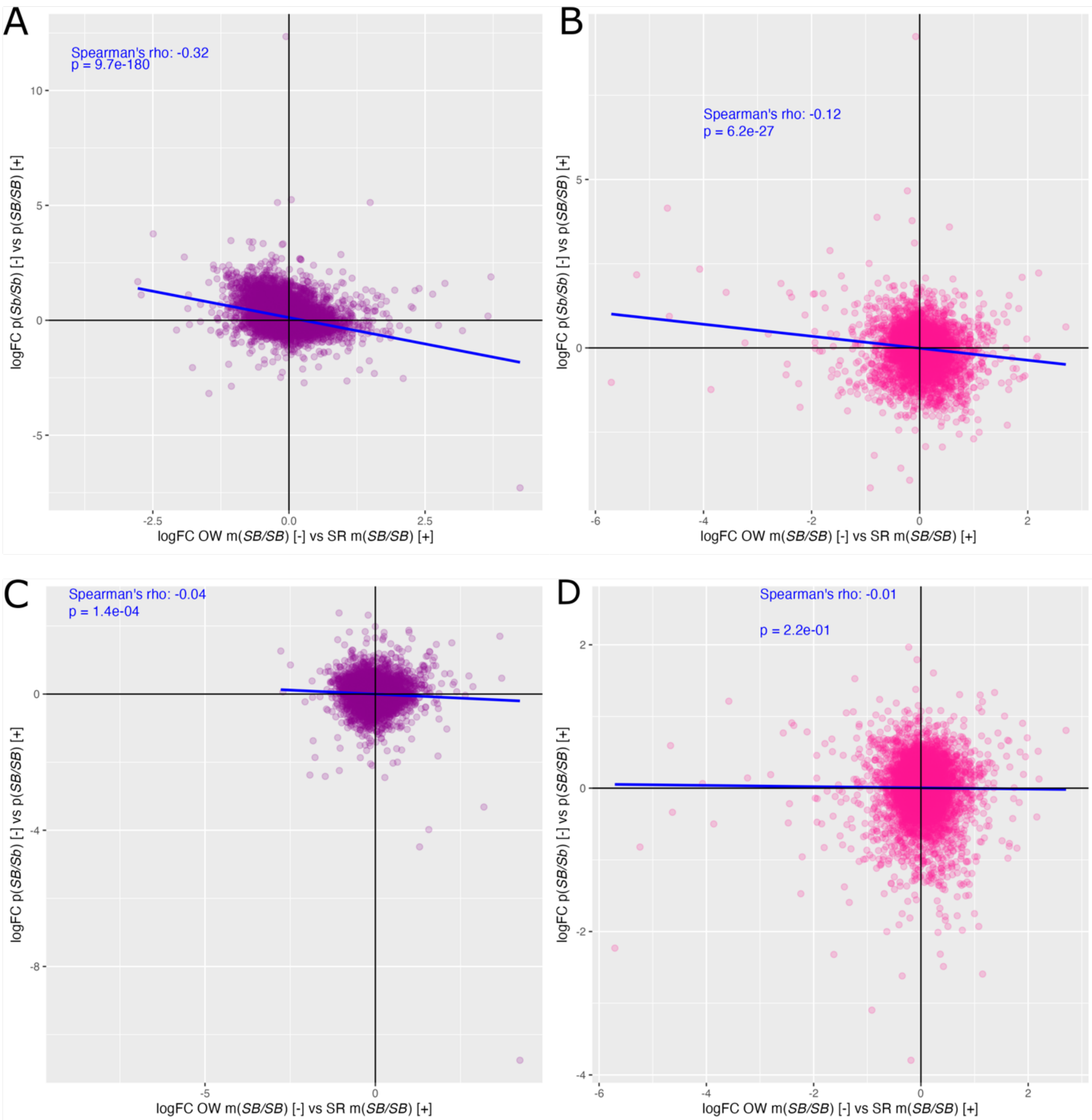


Figure S2.3. Transcriptome-wide correlation between expression fold-change values. Each plot shows log2 fold change (logFC) for two separate gene expression comparisons. For each plot, all genes passing the low expression cutoff for both comparisons are included. The x-axis represents the logFC value for Comparison 1, and the y-axis represents the logFC value for Comparison 2. Plots A and B compare polygyne SB/SB vs. polygyne Sb/Sb gynes, while plots C and D compare polygyne SB/SB vs. polygyne SB/Sb gynes. Brain comparisons are in purple (A & C), and ovary comparisons are in pink (B & D).

Spreadsheets (S1 – S13) can be found as an excel file in supplementary information.

CHAPTER THREE

HOW DO EARLY-CAREER BIOLOGY FACULTY DEVELOP PEDAGOGICAL CONTENT KNOWLEDGE? EXPLORING VARIAITON AND LONGITUDINAL DEVELOPMENT[1]

[1] Waugh, A.H., K.E. Green, and T.C. Andrews. Submitted to *CBE—Life Sciences Education*, 08/16/24.

Abstract

Active-learning instructors are more effective when they use pedagogical content knowledge (PCK) to anticipate, interpret, and respond to student thinking. PCK is topic-specific and includes knowledge of student thinking (e.g., common difficulties) and knowledge of instructional strategies (e.g., effective learning tasks). Currently, we know little about how instructors develop PCK. We documented how 11 early-career undergraduate life science instructors developed PCK over multiple semesters by eliciting knowledge as instructors planned, implemented, and reflected on instruction. Qualitative content analysis indicated that instructors' PCK about student thinking was not necessarily grounded in evidence from students and their PCK about instructional strategies varied in whether and how it considered student thinking. We adapted a rubric to test hypotheses about PCK development trajectories. Participants' PCK about student thinking tended to become more grounded in evidence from students and their PCK about instructional strategies tended to focus more on student thinking over time. However, teaching experience did not necessarily lead to PCK development. Case study analysis revealed that pedagogical knowledge and specific practices supported PCK development. We propose a hypothetical model to explain how teaching knowledge and practices support PCK development. We also suggest reflections and actions for instructors who want to develop their PCK.

Introduction

Teaching an active-learning lesson in undergraduate science, technology, engineering, and mathematics (STEM) courses places unique demands on an instructor compared to traditional, didactic lecture (e.g., Andrews & Lemons, 2015; e.g., Dancy et al., 2016). Whereas a traditional lecturer can plan and deliver lessons without much, if any, deviation from an *a priori* plan, active-learning instructors often ask students to discuss their ideas during class and then must interpret and respond to those ideas in real time (Gehrtz et al., 2022). To meet these unique demands, active-learning instructors draw upon teaching knowledge other than subject matter knowledge to interpret students' thinking and respond appropriately (e.g., Auerbach & Andrews, 2018; Wagner et al., 2007).

One key domain of teaching knowledge for active-learning instruction is pedagogical content knowledge (PCK). PCK is knowledge for teaching a specific topic to students at a particular educational level (Gess-Newsome, 2015). Since PCK is topic-specific knowledge, instructors rely on distinct PCK for each topic they teach (e.g., distinct PCK for teaching natural selection, genetic drift, and phylogenies). PCK has several components, and we focus on two of the most studied (Table 3.1): knowledge of student thinking (KST) and knowledge of instructional strategies (KIS; e.g., Andrews et al., 2022; Chan & Yung, 2015; Depaepe et al., 2013; Park & Oliver, 2008).

KST includes awareness of students' prior knowledge, conceptual difficulties, and common naive ideas related to a topic, and how student thinking about the topic may change with instruction (Magnusson et al., 1999; Park & Oliver, 2008). For example, an instructor's KST for teaching evolution via natural selection could include knowing that students often mistakenly view adaptation as a process by which organisms change within their lifetime to enhance survival and reproduction (Gregory, 2009; Nehm & Reilly, 2007).

Knowledge of instructional strategies (KIS) includes awareness of effective examples, analogies, problems, case studies, visual representations, and other instructional strategies that make a topic accessible to students and facilitate learning (e.g., Magnusson et al., 1999; Shulman, 1986). For example, an instructor's KIS for teaching adaptation could include their awareness of activities or cases that emphasize the distinction between adaptation via natural selection and phenotypic plasticity. An instructor with more developed PCK exhibits a greater diversity of ideas about student thinking and instructional strategies for a given topic, more connections between their KST and KIS for a topic, and clearer focus on core concepts for a topic (e.g., Henze et al., 2008; Sickel & Friedrichsen, 2018).

PCK plays an important role in effective implementation of evidence-based teaching strategies (e.g., Andrews et al., 2022). For example, undergraduate mathematics instructors struggled to implement inquiry-based curricula when they had limited PCK. They could not anticipate the difficulties students would encounter as they engaged in inquiry, nor could they efficiently interpret students' ideas in real-time during class, both of which limited their ability to facilitate student learning (Johnson & Larsen, 2012; Speer & Wagner, 2009; Wagner et al., 2007). In undergraduate biology, more effective active-learning instructors exhibited more PCK, using this knowledge to evaluate, plan, and implement interactive lessons that specifically targeted anticipated student difficulties (Andrews et al., 2019; Auerbach & Andrews, 2018). Similarly, among high school biology instructors, PCK was strongly and positively correlated with the use of evidence-based teaching practices (Park et al., 2011).

Despite the important role of PCK in active-learning instruction, we know little about how instructors develop this specialized teaching expertise. Longitudinal studies of PCK development are scarce, and none have examined undergraduate instructors (Andrews et al., 2022; Chan & Hume, 2019). Though the scope of prior work is limited, it suggests some intriguing areas for further investigation. Teaching experience seems to be necessary but insufficient for PCK development, and pedagogical knowledge can help instructors learn more from their teaching experiences (Chan & Yung, 2018a; Gehrtz et al., 2022; Van Driel et al., 1998).

Pedagogical knowledge is knowledge of teaching and learning that is generalizable, rather than topic-specific (Auerbach & Andrews, 2018). Instructors who exhibit knowledge of monitoring student thinking create more opportunities to access student thinking while teaching, such as through formative assessments, and may develop more PCK and do so more efficiently than other instructors (e.g., Andrews et al., 2019; Auerbach & Andrews, 2018; Chan & Yung, 2018a). Knowledge of how people learn, another type of pedagogical knowledge, may also facilitate PCK development. In a study of three high school teachers over two years, two instructors who demonstrated constructivist ideas about learning more quickly learned about student difficulties with a particular topic (i.e., developed KST) and used what they learned to modify their teaching of the topic (i.e., developed KIS that relied on KST). Relatedly, undergraduate instructors with constructivist ideas designed in-class work that engaged students in generating reasoning, which enabled them to hear student reasoning, creating opportunities to develop PCK (Andrews et al., 2019).

Our research objectives in this study were to characterize variation in PCK among early-career undergraduate life science instructors and examine PCK development over time. To accomplish these objectives, we investigated PCK across multiple semesters, starting in participants' early semesters as college faculty. We used an in-depth data collection approach to elicit the PCK and pedagogical knowledge that participants used as they planned, implemented, and reflected on a lesson, and complemented this with analysis of classroom practices. Here we report the variation in KST and KIS observed among participants, quantified trajectories of PCK development that test emergent hypotheses, and two in-depth case studies that reveal what facilitated an instructor's PCK development.

Table 3.1. Two components of pedagogical content knowledge (PCK) examined in this work, with definitions and examples.

PCK	Definition	Example

component		
Knowledge of student thinking (KST)	An instructor's awareness of students' prior knowledge, conceptual difficulties, and common naive ideas related to a specific topic, and how student thinking about the topic may change with instruction	Awareness of what is easier and more difficult for their students regarding the topic of carrying capacity: *"The biggest thing about carrying capacity that they get wrong is that idea that competitors affect carrying capacity, which is kind of logical, but not how we define it. So, I like that they're being logical about something, and kind of thinking through the idea that a competitor is going to reduce the total number of resources available to an organism. But, that's just not how we define it. So, making that distinction between reality and theory is tricky." - Amy (this study)*
Knowledge of instructional strategies (KIS)	An instructor's awareness of effective examples, analogies, problems, case studies, visual representations, and other instructional strategies that make a specific topic accessible to students and facilitate learning	Awareness of examples that help students learn that carrying capacity is not influenced by the presence of competitors: *"I've added a couple of examples… to kind of drive home this idea that it's resources, that carrying capacity is resource driven not competition or predator driven." – Amy (this study)*

Methods

This work was determined by the Institutional Review Board at the University of Georgia to be exempt (PROJECT00000297).

Participants

We recruited 11 participants who taught large (50+ students) undergraduate life sciences courses and collected data across multiple semesters for each (n= 36 time points) (Table 3.2). Participants self-identified as using active-learning strategies and aiming to continue to develop these strategies. They taught at four public institutions in the Southeastern United States with very high research activity. All participants had earned a Ph.D. in a discipline relevant to the course they taught and held a long-term tenure- (n = 4) or teaching-track (n = 7) faculty position. Participants were early-career faculty who had taught college courses as a faculty member for four or fewer semesters. For each participant, we collected data in one target lesson in one course, and repeated data collection every time they taught the target lesson over two or more semesters (range = 2-5 time points; mean = 3.27; Table 3.2). For some participants, this meant collecting data in the fall and spring semesters and for others it resulted in annual data collection in fall or spring semester. For seven of the 11 participants (64%), the first semester of data collection occurred during their first semester they had ever taught the topics of the target lesson.

Participants taught a variety of courses, including upper- and lower-division life science courses in biology, microbiology, developmental biology, evolutionary biology, and neuroscience, to majors and mixes of majors and non-majors. Participants' focal courses enrolled an average of 135 students (SD = 64.5; Table 3.2). Data collection occurred Fall 2019 through Spring 2023 semester, excluding summers, with different durations for different participants, depending on when they joined the study and when they stopped teaching the focal course (Table 3.2). Due to the COVID-19 pandemic, some semesters of data collection involved remote teaching or HyFlex formats (Figure S1).

Participants stood out from typical life sciences faculty in their level of engagement in teaching development. At the end of the study, we surveyed participants about their teaching professional development and related experiences from both before and during their participation in our study. We learned that most (8 of 11 participants) had participated in 40+ hours of teaching professional development and had formal training in teaching as a graduate student or postdoctoral researcher (9 of 11 participants). Roughly half (6 of 11) had engaged in a formal teaching mentorship as a mentee, and some (3 of 11) had, themselves, led teaching professional development. Two participants had published peer-reviewed discipline-based education research, and another two had published biology course materials or lessons. This high level of engagement in teaching and education-related work likely meant that participants had motivations, knowledge, and skills that set them apart from other life sciences faculty.

Table 3.2. Participant information.

Pseudonym	# of time points	Course level [a]	Approx. class size	Position	Semesters teaching the focal topic at first time point	Semesters teaching experience as faculty at first time point
Amy	5	LD	180	Lecturer	0	4
Beth	4	UD	80	Assistant Professor	0	0
Claire	4	LD	100	Lecturer	0	0
Dana	4	UD	110	Assistant Professor	2	2
Eric	4	LD	60	Academic professional	2	2
Fiona	3	LD	75	Academic professional	0	1
George	3	LD	250	Lecturer	1	2
Henry	3	LD	200	Lecturer	0	4
Irene	2	UD	200	Assistant Professor	2	2
June	2	UD	150	Lecturer	0	2
Kim	2	LD	75	Assistant Professor	0	0

[a] UD = upper division; LD = lower division

Operationalization of PCK

In this study, we operationalize PCK according to the most contemporary theoretical framework, the reformed consensus model of PCK (Carlson et al., 2019). PCK is closely linked to action and includes both skills and reasoning used in planning, enacting, and reflecting on instruction (Gess-Newsome, 2015). PCK encompasses both explicit knowledge, which an instructor can readily articulate, and tacit knowledge and skills, which could be challenging for the instructor to express , and may be drawn on heavily in real-time while making instructional decisions (Alonzo et al., 2019; Alonzo & Kim, 2016). We studied two components of PCK (KST and KIS, Table 3.1) because these are the most studied components of PCK and are considered crucial in shaping an instructor's overall PCK (Andrews et al., 2022; Chan & Yung, 2015; Depaepe et al., 2013; Park & Chen, 2012; Park & Oliver, 2008). PCK is specific to each topic that an instructor teaches at each educational level (Chan & Hume, 2019). Given that life science education prioritizes both core concepts and competencies (AAAS, 2011), instructors likely need PCK about each topic and each competency that they aim for their students to learn. Therefore, we defined PCK for undergraduate life sciences as including both ideas and rationales related to biology topics and biology competencies, as defined in Vision & Change (AAAS, 2011; Clemmons et al., 2020). Our operationalization of PCK led us to a methodology that captured PCK: (1) used in planning, enacting, and reflecting on instruction, (2) in the authentic context of an instructor's classroom and teaching practice, and (3) by engaging instructors in reflection about their real-time reasoning while teaching (Alonzo et al., 2019).

Eliciting pedagogical and pedagogical content knowledge

We conducted pre- and post-lesson interviews at each time point to elicit the PCK and pedagogical knowledge that participants used as they planned, implemented, and reflected on a target lesson, following the methods of Alonzo et al. (2019). The interviews aimed to elicit knowledge closely linked to the target lesson, including knowledge of student thinking (KST) and knowledge of instructional strategies (KIS) and generalizable pedagogical knowledge about monitoring student thinking and how people learn (e.g., Andrews et al., 2019; Chan & Yung, 2018a). We conducted a pre-lesson interview with participants before they taught the target lesson. Then, we video-recorded the target lesson and created short video clips for a post-lesson interview that used stimulated recall.

Pre-lesson interviews aimed to elicit PCK and pedagogical knowledge used in planning the target lesson. We obtained lesson materials (slides, handouts, pre-class work, etc.) before the pre-lesson interview. Using these materials, we began the semi-structured interviews by probing the rationale behind the design of lesson materials and, if applicable, any changes to the lesson materials since the previous time point. Questions in the pre-lesson interview targeted participants' PCK by asking, for example, about their awareness of students' prior knowledge related to the lesson topic(s) (KST), any anticipated difficulties with those topics (KST), and how their chosen instructional strategies would help students overcome those difficulties (KIS). Pre-lesson interviews also targeted pedagogical knowledge by asking participants how they know what students understand and when they are confused about a topic during class and how a particular instructional strategy used in the target lesson helps students learn (see full interview protocol in Appendices A & B). Pre-lesson interviews typically lasted ~60 minutes and occurred no more than one business day before the lesson.

We video-recorded the target lesson using a camera in the rear of the room and a lapel microphone worn by the instructor. These video-recordings provided: (1) video clips for the post-lesson interview, and (2) data for systematic analysis of teaching strategies (described below). Prior to each post-lesson interview, we used a set of criteria to select three or more clips (range = 3-6 clips) from the video-recorded lesson. First, to facilitate our ability to elicit participant's PCK, we prioritized lesson segments where the instructor accessed student thinking by talking to students, seeing student work (including work on multiple choice "clicker" questions using classroom response systems), and hearing student questions. Since instructors often accessed student thinking at multiple points in the lesson, we secondarily prioritized lesson segments that showed students engaged in generative cognitive work (Andrews et al., 2019; Chi & Wylie, 2014) Generative cognitive work involves students working individually or collaboratively to generate ideas and products that go beyond what has been presented to them (Chi & Wylie, 2014). When necessary, we tertiarily included contrasting lesson segments that showed students engaged in "active" cognitive work, which involves recall or algorithmic problem-solving.

Post-lesson interviews aimed to facilitate participants' ability to recall the knowledge they used in real-time while teaching and to elicit knowledge drawn upon in reflecting on the lesson (Alonzo & Kim, 2016). In each post-lesson interview, the participant and interviewer watched selected video clips of the recorded lesson. We asked participants to provide a "running commentary" of everything they remembered thinking during the teaching moments depicted in the clips. After participants shared their memories, we asked follow-up questions to further probe participant's thinking. The protocol for these questions had to be adaptable, rather than semi-structured, because the content of the clips depended on the instructor's lesson. We built a protocol for each interview by selecting questions from a pool of questions that best fit the lesson. To ensure that each interview thoroughly elicited both PCK and pedagogical knowledge, each protocol included a minimum number of questions focused on PCK and a minimum focused on pedagogical knowledge. One question used to elicit KST asked, "Can you walk me through what you were thinking about this student's reasoning about [specific topic]?" A question used to elicit KIS asked participants, "In what way is this [specific question] particularly useful in helping students learn [specific topic]?" (see full interview protocol in Supplemental Materials Appendices A & B).

Qualitative analysis of teacher knowledge

We aimed to identify PCK and pedagogical knowledge used by participants as they planned, implemented, and reflected on a target lesson and to characterize features of this PCK that varied across and within participants. We first analyzed interview transcripts for evidence of PCK using two *a priori* codes corresponding to two core components of PCK: knowledge of student thinking (KST) and knowledge of instructional strategies (KIS; Andrews et al., 2019; Auerbach & Andrews, 2018; Magnusson et al., 1999; Park & Chen, 2012; Park & Oliver, 2008). We also analyzed transcripts for evidence of pedagogical knowledge with two *a priori* codes corresponding to two categories of pedagogical knowledge that prior work suggests may influence PCK development, knowledge of monitoring student thinking and knowledge of how people learn (Andrews et al., 2019; Auerbach & Andrews, 2018; Chan & Yung, 2018b). Knowledge of monitoring student thinking includes an instructor's awareness of approaches to elicit detailed student thinking during class and the purposes of accessing student thinking and altering instruction accordingly. Knowledge of how people learn includes an instructor's awareness that students learn from engaging in generative work during class, which is work that requires students to generate something beyond what has been presented to them, knowing the types of tasks and problems that require generative work (Chi & Wylie, 2014), and understanding how to facilitate student work to maintain generative cognitive engagement (e.g., Andrews et al., 2019; Auerbach & Andrews, 2018).

As we accumulated examples of each of these types of PCK and pedagogical knowledge, we could characterize variation. We used constant comparison to create new codes, nested within the four original *a priori* codes, to distinguish knowledge features. At least two researchers (A.H.W., K.E.G., and T.C.A.) independently read and coded each transcript and then we discussed all disagreements to reach consensus. As we refined the definitions and bounds of codes, we re-analyzed all previously coded data. The outcome of this work was a finalized codebook (Supplementary Table S3.1) We developed the codebook in the first years of the project, using about 50% of the data, and then applied it to the full set of interviews across time for each participant.

The variation in knowledge exhibited by participants, along with prior research, allowed us to generate hypotheses about trajectories of PCK development before we had collected the full longitudinal dataset. We share brief descriptions of that variation here, as subsequent methods follow from these early findings. Participants' KST varied in whether it was based on evidence of their students' thinking. We refer to KST grounded in evidence of student's thinking as "student-validated" KST. Participants' KIS varied in the extent to which it was based on knowledge of student thinking, which is referred to as integrated KIS (Park & Chen, 2012). We describe these findings in detail in the results.

We hypothesized that early-career life sciences faculty would develop increasingly student-validated and integrated PCK as they gained experience teaching the same topic(s) over time. To test these hypotheses within our sample, we needed a way to systematically assess changes in the extent of student-validation and integration of participants' PCK across time. This required reducing our rich qualitative data in ways that could be compared across time. We achieved this by developing a PCK rubric that assigned numeric scores to represent evidence of PCK (Table S3.2). To do this, we drew on elements of an existing PCK rubric (Park & Oliver, 2008; Saldaña, 2003).

Our PCK rubric included four categories (i.e., rows) with four levels (i.e., columns). The four categories captured KST and KIS used in planning the lesson (pre-lesson interview) and in implementing and reflecting on the lesson (post-lesson interview). Increasing levels of KST in the rubric corresponded to an increase in the number of unique topics for which an instructor demonstrated student-validated KST. Increasing levels of KIS in the rubric corresponded to an increase in the number of unique topic-specific instructional strategies for which an instructor demonstrated integrated KIS and increasing use of instructional strategies tailored to a particular student difficulty rather than strategies that were more general in scope.

We analyzed the PCK exhibited by a participant at each time point without awareness of the scores at other time points. Two researchers (AW, KG, or TC) independently determined rubric scores, using coded transcripts. We used a quadratic weighted Cohen's kappa to determine inter-rater reliability (IRR) for rubric scores. Cohen's kappa weights disagreements according to their squared distance from perfect agreement. This approach penalizes disagreements with increasing severity as ratings fall farther apart from one another on an ordinal scale. To calculate IRR, we treated each of the four categories (rows) of the rubric as a separate judgment, resulting in four judgements per participant per time point. We achieved high inter-rater reliability (IRR; quadratic weighted Cohen's kappa = 0.835; 144 judgements) and discussed all disagreements to consensus. We plotted PCK rubric scores over time using line plots depicting the contribution of the KST, KIS, and total PCK rubric scores (Figure 3.1).

Case study selection and analysis

Quantifying qualitative data necessarily reduces the complexity and overlooks important nuances in why and how participants developed PCK over time. We aimed to uncover these nuances by developing detailed case studies of a few participants. We selected participants for whom we had captured all their experience teaching the focal lesson (i.e., first semester of data collection coincided with their first semester teaching the focal lesson) and who had robust PCK growth, quantified as an increase of two or more points on the PCK rubric every semester. After applying these criteria, three participants remained: Fiona, Henry, and June. For each, we compiled all evidence of their PCK from interviews and teaching materials. After compiling evidence of PCK development for each of these participants and writing initial descriptions of their PCK development, we selected Henry and June as case study subjects based on illuminating patterns of PCK development as well as the overall clarity of the participant's quotations for the purposes of reporting.

We continued to refine our case descriptions of Henry's and June's PCK development until two researchers (A.H.W, T.C.A) agreed that the case description was complete, accurate, and accessible to a wide audience. We then returned to each participant's full corpus of data (interviews, classroom recordings, and teaching materials) to further identify factors that supported their PCK development. The two researchers independently generated a list of teaching practices and pedagogical knowledge components that contributed to each participant's PCK development and discussed them to reach consensus. Then, we wrote a synopsis of each participant's PCK development and the pedagogical knowledge and practices that supported it over the longitude of the study. We also independently generated visual representations of the supporting knowledge and practices before discussing these representations to arrive at a consensus visualization for each participant. The outcome of these analyses for each case includes a summary of context and practices, a detailed description of PCK development with supporting quotes, and a synopsis of the thinking and actions that facilitated the participant's PCK development.

Trustworthiness of these qualitative analytic approaches

Trustworthiness in qualitative data analysis involves four domains: credibility, confirmability, dependability, and transferability (Anfara et al., 2002; Guba, 1981; Shenton, 2004). In this study, we intentionally incorporated data collection, analytic, and reporting practices to achieve a trustworthy qualitative approach.

Credibility in qualitative research is achieved by aligning the planned focus of the study with the research methodology so that the investigation measures what is intended (Shenton, 2004). Our data are credible because we used interview approaches grounded in the existing literature and that precisely addressed our question about PCK development (Alonzo & Kim, 2016; Park & Chen, 2012). We utilized well-established methods to access teacher knowledge, including stimulated recall (Alonzo & Kim, 2016), to obtain evidence of PCK, thereby ensuring our ability to accurately capture and measure the intended construct. Our data collection involved interview questions that drew on prior studies of PCK in a similar context (Andrews et al., 2019) to provide access to the PCK that early-career instructors use in planning, enacting, and reflecting on instruction.

Confirmability and dependability in qualitative research are achieved by ensuring that interpretations of qualitative data are consistent with the original data (confirmability) and repeatable by different researchers (dependability; Anfara et al., 2002; Shenton, 2004). Our use of constant comparison enhances the confirmability and dependability of our qualitative analysis. Multiple researchers immersed themselves thoroughly in all the data, ensuring that interpretations were not the responsibility of any one individual. In both our qualitative content analysis and case studies, our methods involved regular independent analysis and collaborative discussions to refine ideas. We emphasized collective idea generation and mitigated power imbalances between researchers. By requiring consensus, we prevented any single viewpoint from dominating the data interpretation.

Transferability in qualitative research benefits from clarifying the extent to which the results of a study might be generalized or transferred to other contexts (Shenton, 2004). We improve the overall transferability of this study by reporting relevant details about our participants, their teaching contexts, their institution type, and other factors that could influence their PCK development. We also use careful language to limit all findings and claims to the study population as we lacked the sample size and diversity of participants to propose generalizable patterns of variation and PCK development.

Analyzing teaching strategies

We systematically analyzed teaching practices used by participants during each target lesson. We analyzed videos to quantify the use of three specific behaviors: instructors working with students in small groups or one-on-one, one or more students asking a question and/or sharing an answer verbally with the whole class, and students explaining their reasoning to the instructor. The first two behaviors were taken directly from the Classroom Observation Protocol for Undergrad STEM (Smith et al., 2013). We developed the third because we aimed to quantify the access that instructors sought to hear students explain their thinking.

We defined students explaining their reasoning to the instructor as any instance where a student verbally explains or elaborates on their thinking to the instructor. This could have occurred when a student explained their thinking one-on-one, in small groups, or to the whole class, as long as the instructor was listening. We required that the students' answers be more than just a word or phrase, to limit inclusion to elaborations on student thinking.

We focused on these classroom behaviors because they place instructors in contact with evidence of students' thinking during class and, consequently, may contribute to PCK development (e.g., Chan & Yung, 2018a; Gehrtz et al., 2022). We quantified these student and instructor behaviors by marking the presence or absence of each behavior in each two-minute segment of the video-recorded class and calculating the average percent of two-minute segments including each behavior across target lessons for the instructors examined in cases. A coding team was trained on this protocol as part of a larger project and determined inter-rater reliability (IRR) by calculating Cohen's kappa. After establishing high IRR among coders, individual coders analyzed remaining videos, calculating their IRR for every fifth video to ensure high IRR across coders and time. Out of 54 videos analyzed for this study, multiple coders analyzed 20, with an average Cohen's kappa of 0.96.

Results

We present results in three sections, which build upon each other to address our research objectives. The first section describes qualitative variation in the two components of PCK examined in this study, knowledge of student thinking (KST) and knowledge of instructional strategies (KIS) and draws on the full dataset. Based on the variation observed early in our longitudinal data collection, we hypothesized about how PCK would develop among participants. The second section presents a longitudinal quantification of PCK development that tests our hypotheses. The quantitative data help reveal patterns over time and across participants, but also necessarily reduces much of the important complexity in how participants' knowledge developed. Therefore, the third section presents in-depth case studies of PCK development in two participants over time. All text within quotations are verbatim quotes from participants, with minor editing for clarity.

How did PCK vary among participants?

In this section, we provide quotations from pre- and post-lesson interviews to illustrate the qualitative variation in KST and KIS used by participants in their teaching.

Variation in knowledge of student thinking (KST)

The knowledge of student thinking that participants used in planning, implementing, and reflecting on their target lesson varied in whether it was grounded in evidence from their students (Table 3.3). Participants who had previously elicited and paid attention to student thinking about a topic could demonstrate KST grounded in evidence about student thinking, which we refer to as student-validated KST (Table 3.3). Participants had made these observations in previous semesters or related courses, through student thinking displayed in exams, formative assessments before or during class, and in conversations with students. For example, when reflecting on a lesson on mammalian sex determination, Dana explained that they had observed how students readily grasp the mechanics of primary sex determination because it is “a little more black and white” to understand how XX and XY individuals develop distinct traits, but that students struggle with secondary sex determination because it hinges on relative hormone levels and other complex mechanisms.

Though KST, by definition, is about how students think and learn about a topic, often participants did not rely on evidence from students to inform their KST. Instead, participants made assumptions based on their own thinking and/or experiences, made guesses based on their perceptions of thinking among the general public, or presumed that students would lack any relevant ideas about a topic (Table 3.3).

Since we deliberately studied participants new to teaching, it makes sense that they might lack KST grounded in evidence from students, yet not all participants replaced guesses and assumptions about student thinking with student-validated KST as they gained experience (Table 3.3). One way that participants developed student-validated KST involved treating their assumptions about student thinking as hypotheses and testing them against observations of student thinking. For instance, Claire anticipated that their neuroscience students would think that dopamine directly causes the pleasurable feelings associated with a stimulus, rather than motivating behavior toward obtaining that reward, because they thought many members of the general public hold this idea. Claire anticipated this naive idea across the first three semesters of their involvement in this study, but they remained uncertain about how many students actually held that idea. In their fourth semester, Claire created a clicker question that would elicit the idea, and explained their rationale this way, “Over time, I've been saying, ‘Oh, from what I understand this is what students think dopamine is mainly involved in.’ And so, then I said, ‘You know what? I should actually assess this instead of making some assumptions from some students that I've heard from.’ So, I finally made it an actual clicker, and the results were pretty much what we had expected… I wanted to see if there was any evidence to this kind of hunch that I had about their initial ideas.”

Table 3.3. Variation in knowledge of student thinking (KST), with descriptions and example quotes. Lesson topic in parentheses following each quote.

Knowledge description	Example quotes
Student-validated KST: This knowledge relied on observations of students’ thinking about a topic, including (but not limited to) evidence from exams, formative assessments,	*Beth observed a pattern of student thinking in an assignment students completed before class:* “I only have a small handful of the pre-assessments… A lot of students think about pluripotent stem cells, meaning the ones that can become all the types of the adult body. So when they

and interactions with students in and out of class.	answer 'what is a stem cell?,' they almost all answer that, not that there are different kinds." (stem cells)
	Irene observed a pattern of student thinking during class: "I know last time [I taught the lesson] they were definitely having a problem with the concept that plasticity is change within a generation, yet evolution is change between generations and so sort of just getting confused as to what role phenotypic plasticity has played." (phenotypic plasticity)
KST lacking evidence from students: This knowledge relied on assumptions and guesses, rather than observations or information from students, or lacked evidence of an origin.	*George made a guess about student thinking based on perceptions of the general public:* "My understanding is students have a vague idea of what ecology is. They've probably seen stuff on TV, Animal Planet, National Geographic, whatever, which are folks going out and looking at animals. You know, the naturalist with the binoculars in the tree stand. […] What they might miss is what else is going on between that animal and its bigger population and its environment." (carrying capacity)
	Eric made an assumption about student thinking based on their own experience: **"**I don't know if they have many ideas. I don't know how many of these concepts are really introduced. I mean, in terms of broad tuning and fine tuning, the misconception may be that a neuron fires because it fires, right? Regardless of when and what signal is. [...] I don't know if [they have that idea] because of misinformation, but more likely just not having haven't been told before. I have no idea, honestly, how much the high school biology curriculum has changed since I went through, but I don't remember ever being taught differences in cell firing as contributing to sensing things in the environment. So, this could be a pretty new principle for some." (combinatorial coding in neurons)

	Fiona's idea about student thinking did not have a clear origin: "I would predict that they're going to have the most trouble with the fact that it's a probabilistic relationship. They may not have that nuance of their thinking yet, and I'll try to scaffold them through that." (motivated behavior)

Variation in knowledge of instructional strategies (KIS)

The knowledge of instructional strategies that participants used in planning, implementing, and reflecting on the target lesson varied in the extent to which it relied on knowledge of student thinking (Table 3.4). Sometimes participants displayed integrated KIS, which drew on their KST for the same topic, whereas other instances of KIS did not rely on any KST. In other words, sometimes an instructor's rationale for a topic-specific instructional strategy was based on their knowledge of how students tend to think about the topic and other times it was not. We also observed two types of integrated KIS: tailored and broad strokes.

Participants using tailored KIS explained their development or adoption of instructional strategies that they had carefully designed to address the specific difficulties they anticipated students would encounter in learning a topic (Table 3.4). Participants using tailored KIS often designed opportunities for students to encounter an anticipated difficulty in a productive manner. For example, Beth explained that their students tend to jump straight into forming a biological interpretation when reading a particular graph in their lesson, skipping the crucial step of describing patterns in data. Beth exhibited tailored KIS when they explained how they aimed to address this pattern of student thinking by splitting the question about the graph into two parts, asking students to describe the data before synthesizing an interpretation.

Not all participants who displayed integrated KIS had such tailored strategies. Participants sometimes drew on KST in their reasoning for an instructional strategy that might be useful for a wide range of difficulties that students encounter. We employ the phrase "broad-strokes KIS” to describe this knowledge. The term “broad strokes” refers to something that is more general than it is detailed (big strokes with a paint brush, rather than detailed brush work). Broad-strokes KIS was integrated with KST, but the instructional strategy did not address the details of a student difficulty. These instructional strategies might involve adding more examples of a topic, providing more explanation, or offering additional practice with the topic (Table 3.4). For example, Fiona described how students struggle with interpreting one graph that had two different y-axes, so they aimed to “walk them through” the graph so it was clear what was depicted on each y-axis. For both tailored and broad-strokes KIS, the knowledge of student thinking might be student validated or lack evidence from students.

Commonly, participants exhibited KIS lacking any evidence of a basis in KST. The reasoning behind these instructional strategies focused on other priorities besides the ways in which students would think about and learn a topic. Primarily, this kind of reasoning consisted of topic-specific instructional decisions based on content priorities. For example, Eric explained how an activity in their neuroscience lesson on olfaction was useful because it demonstrated how some olfactory neurons respond to a wide range of stimuli, whereas others are finely tuned to a specific stimulus. At times, this kind of KIS included reasoning focused on logistics (e.g., timing), ideas about students’ characteristics (e.g., first-year vs fourth-year, rural vs. urban, more vs. less engaged), or the basis of the instructional strategy rationale was unclear (Table 3.4).

Table 3.4. Variation in knowledge of instructional strategies (KIS), with descriptions and example quotes. Lesson topic in parentheses following quotes.

Knowledge description	Example quote

Tailored KIS: This knowledge relied on KST[1] as the rationale for specifically addressing an anticipated student difficulty with the topic.	*Irene describes the reasoning behind the design of a multiple-choice question posed to students in class:* “I deliberately have picked here the difficult [answer choices], the ones that they do confuse, which is genetics by environment and G[enetics] plus environment. So, I think that's potentially going to be hard.” (genotype-environment interaction)
	Dana explains their rationale behind a prompting question they posed as they talked with a small group: “It was that little detail of well okay so yes, embryo is not making anything. We all agree on that, but is there something else that can feed into this? So, I think that that piece, that is one where, you know, for some students the second I say, ‘Well, this embryo is human. We're developing within a maternal environment, you can just see the light bulb and they're just like, ‘Oh. And that's why that is.’” (sex determination)
Broad-strokes KIS: This knowledge relied on KST as the rationale for a general instructional strategy that might be useful for a range of difficulties students might face.	*Claire explains their rationale for slowing down and spending more time explaining a topic:* “[Knowing students have a misconception about dopamine] makes me slow down because then I can realize like, ‘Okay, this is what we need to spend more time on, and that's why I'm going to spend that one slide from the last lecture, to go over just a little bit more of that’. But yeah, so I spend more time explaining, and I try to come up with different examples or different ways to explain it if my original explanation maybe wasn't as clear or didn't really resonate with them.” (dopamine and motivation)
	Amy explains their rationale for keeping things simple, using examples, and gradually introducing complexity. “It's kind of a balance between wanting to be transparent about the uncertainty around this idea and like the ambiguity around carrying capacity, while also simplifying it to a level that's applicable across some number of examples. So, I think I try to keep it simple. I'll teach it as a simple concept and straightforward without too much nuance initially, and then kind of add in at the end that, ‘This is the simplest view of this that we could have. There's lots of complicating factors.’” (carrying capacity)
KIS lacking evidence of integration: This knowledge lacked	

evidence of a link to KST, often focusing on content, student characteristics, teaching context, and other factors.	*Kim explains that they removed details from the lesson because students were less engaged and struggled with complex material compared to the prior semester:* “This year I feel like my class is kind of struggling. They're not able to just kind of immediately grasp sort of complex ideas in the way that my class last year was. So, I have been cutting some things out that don't seem as important. I was looking at those slides about carbon isotopes and thinking, ‘It probably will take me 10 minutes to explain this in a way that folks in the class will understand, and it's not really worth 10 minutes of instructional time.” (carbon dating to measure water use efficiency)
	George explains that they will demonstrate mark and recapture to help students imagine ecological methods: “I think it's good cause we were talking, cause one of the big questions was, ‘If you're an ecologist, how do we actually study a population if we can't actually count it?’ I think understanding one method that they could use… that's actually pretty simple and hands-on, it might get them a little perspective on how, you know, it's not just a bunch of biologists and they're going, ‘I think there's about 300 individuals here.’” (population size estimation)

Longitudinal quantification of PCK development

Based on the variation in KST and KIS observed in data collected early in our study, as well as prior research, we hypothesized that participants’ PCK would develop in two specific ways as participants gained teaching experience:

Hypothesis 1: Participants would display more KST that was grounded in evidence from students (i.e., student validated).

Hypothesis 2: Participants would display more KIS that was based on KST (i.e., integrated; Park & Chen, 2012; Sickel & Friedrichsen, 2018).

To test these hypotheses, we quantified the PCK used by participants in their planning, implementing, and reflecting on the target lesson and plotted these quantifications over time (Figure 3.1). A PCK rubric score of two indicates a lack of student-validated KST (red line) and a lack of integrated KIS (blue line) and a score of eight indicates the participant displayed five or more distinct student-validated ideas (i.e., student-validated KST) and displayed rationales for three or more distinct instructional strategies tailored to student thinking (i.e., tailored KIS) in each interview.

The PCK trajectories supported the hypotheses for most, but not every participant. Every participant had a higher total PCK score at the final time point compared to their first time point (Figure 3.1), indicating that experience corresponded to PCK development in the hypothesized ways. When we look more closely at the components of PCK, seven participants (64%) had higher KST and KIS scores from their first to last data collection (Figure 3.1). The remaining four participants developed one but not both components in the hypothesized ways.

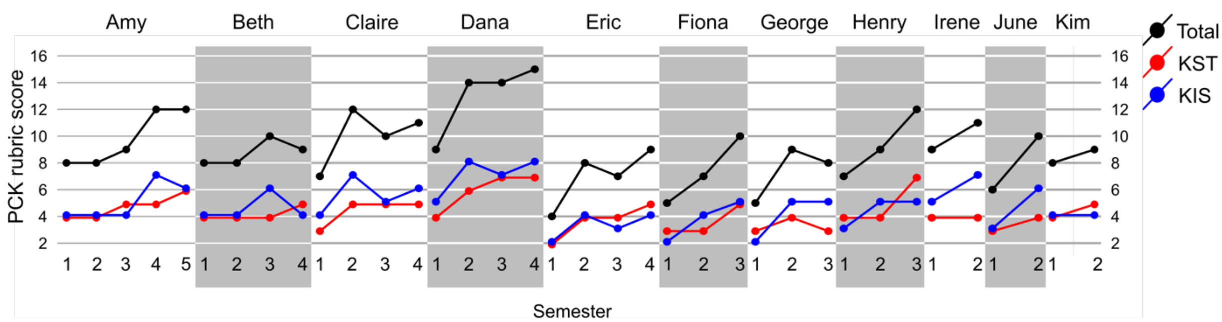


Figure 3.1: Longitudinal trajectories of PCK rubric scores, by participant. KST = Knowledge of student thinking; KIS = Knowledge of instructional strategies. Total = KST + KIS scores. PCK scores combine data from the pre- and post-lesson interviews. Data include each semester an instructor taught the target lesson over a two- to four-year period. Min and max scores are two and eight, respectively, for KST and KIS, and twice that for Total.

Participants tended to develop more student-validated KST with experience teaching the same lesson repeatedly but were not necessarily learning from each iteration of teaching the lesson (Hypothesis 1; Figure 3.1). Most participants (9/11) exhibited student-validated KST about a greater number of topics at the end of the study than at the beginning (Figure 3.1), meaning they had developed awareness of student thinking based on observations of their students. When we examine change between one semester and the next (i.e., sequential time points) for each participant, KST score increased in 52% (13/25) of sequential time points, remained the same in 44% (11/25), and decreased in 4% (1/25). This indicates that participants were not consistently gleaning new insights about student thinking each time they taught the lesson, even though they were quite new to teaching the topic. It also indicates that new student-validated PCK tended to be stable over time, as decreasing scores from time point to time point were rare.

Most participants (9/11) displayed more integrated and tailored KIS with experience teaching the same lesson repeatedly but were not necessarily learning from each semester of teaching (Hypothesis 2; Figure 3.1). These quantifications mean that participants tended to reason about topic-specific instructional strategies with more reliance on their knowledge of student thinking as they gained experience. KIS score increased in 56% (14/25) of sequential time points, did not change in 24% (6/25), and decreased in 20% (5/25). This indicates that, more often than not, participants were relying on more instructional rationales grounded in KST compared to the prior semester. Other times, participants had similar KIS as the prior semester. In almost a quarter of cases, participants demonstrated fewer instances of integrated KIS compared to the prior semester, indicating this knowledge was less stable than KST in these participants. In summary, participants' KST generally became more student validated and their KIS more integrated with additional teaching experience with their lesson, but participants did not necessarily develop PCK each semester.

PCK development case studies

We developed two case studies to examine nuances in why and how participants developed PCK over time. We present each case in three parts, which collectively illustrate how the participant developed PCK for a specific topic: (1) target lesson context, determined from teaching materials and systematic analysis of teaching strategies; (2) a description of their PCK development**,** which includes the KST and KIS the participant displayed each semester, illustrated with quotes, for one topic within their target lesson, as well as evidence factors influencing PCK development; (3) a synopsis of the thinking and actions that facilitated the participant's PCK development, with accompanying visual representation.

PCK development case one: Henry teaches epistasis

This case documents how Henry developed PCK for teaching the topic of epistasis across three semesters of an introductory biology course, starting with his first term teaching the topic.

Target lesson context: Henry's lesson information and teaching practices

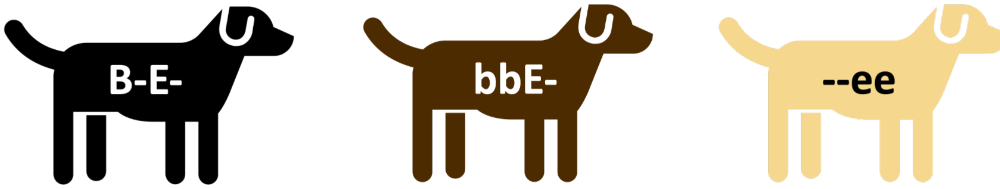
Though new to teaching epistasis, Henry had taught the same introductory course for three prior years, typically teaching three sections of ~150 students per semester in 50-minute class periods. Due to the COVID-19 pandemic, Henry's teaching modality shifted during the study from synchronous in Zoom, to in-person with alternating halves of the class present, to fully in-person. This limited his access to student thinking in the first semester that we studied.

Henry taught epistasis using an in-class activity within a class period about exceptions to Mendelian inheritance within a unit about genetics. The lesson also addressed sex-linked traits, incomplete dominance, and quantitative genetics, though Henry ultimately cut the latter topic to allow more time for the others in the third semester. Epistasis refers to the interaction between two or more genes, such that the expression of one gene masks or modifies the expression of another gene. To teach this topic, Henry first explained a commonly used example of epistasis in which two genes interact to determine coat color in Labrador retrievers (black, brown, or yellow). Alleles of the B gene result in cells producing black (B) or brown (b) pigment, whereas alleles of the E gene result in successful deposition of pigment into fur (E) or failure in pigment deposition (e). Individuals that are homozygous recessive for the E gene (*e/e*) do not successfully deposit pigment into the fur and are yellow regardless of genotype for the B gene.

After this explanation, Henry asked students to predict the phenotypic ratio among the offspring of a cross between *BbEe* and *bbEe* individuals in a multiple-choice question (Figure 3.2). In the third semester, Henry added a second multiple-choice question related to this same example of epistasis, which required students to use a different problem-solving approach to determine the most likely parental genotypes, given a particular offspring phenotypic ratio (Figure 3.2).

Students in Henry’s class spent much of class working on questions or problems independently and in small groups, which Henry referred to as cooperative learning. While students worked, Henry circulated the classroom. In semesters two and three (in-person semesters), analysis of Henry's classroom video recordings showed that students were working on in-class activities for 45% and 70% of the two-minute segments, respectively. Additionally, Henry had one-on-one interactions with students in 37% and 59% of two-minute segments, and students explained their thinking in 13% and 16% of the segments, respectively. Compared to the average target lesson in our sample, Henry spent nearly twice as much class time interacting one-on-one or in small groups (mean: 48% vs 25% of two-minute segments) and nearly twice as much time hearing students explain their thinking as other participants (mean: 15% versus 8% of two-minute segments).

One gene codes for a protein that is needed to produce the pigment, and black pigment (B) is dominant to the chocolate brown (b).



Another gene encodes a protein involved in depositing the pigment in hair follicles, and pigment (E) is dominant to no pigment (e).

1. If male and female Labrador retrievers with genotypes BbEe and bbEe are crossed, what phenotypes are expected for their offspring?

A. 3/8 black, 3/8 brown, 1/4 yellow

B. 1/4 black, 1/4 brown, 1/2 yellow

C. 3/8 black, 1/8 brown, 1/2 yellow

D. 1/2 black, 1/4 brown, 1/4 yellow

(Sem1, Sem2, & Sem3)

2. If a yellow lab and a chocolate lab have several litters with 17 black and 18 chocolate offspring, what are the probable genotypes of the parents?

A. BbEe and bbEe

B. Bbee and bbEE

C. BBee and bbEe

D. not enough info

(Sem3 only)

Figure 3.2: Henry's epistasis activity. Information given to students (top) and multiple-choice questions administered to students via audience response system after small group work (bottom). Correct answers are provided for the reader in green text but were not shown to students until debrief. Henry used Question 1 in semesters one, two, and three and Question 2 in semester three only. Henry generated these questions. The Labrador image was adapted by Henry for class and by us for this publication from "LabRetColors.png" by Elizabeth Arellano, 2007 https://en.wikipedia.org/wiki/File:LabRetColors.png.

PCK development summary: Henry's PCK development for teaching epistasis

In the first semester of data collection, Henry did not display extensive student-validated KST or integrated KIS for teaching epistasis, which is reflected in relatively low PCK scores (Figure 3.1, see semester 1). Henry explained that he did not have "preset ideas of what students are going to struggle with or not struggle with." When reflecting after the lesson, Henry explained, "everything that's happening now is just like me, real-time, getting data on what students are having trouble with." Due to the virtual format in his first semester teaching epistasis, Henry interacted directly with only a small proportion of his students. He reflected that he had "a pretty good idea of how a small subset of the students think about [epistasis]. I guess my biggest question would be 'how does the broader group think about these topics?'" When asked why hearing from the broader group was more advantageous, Henry said, "I mean isn't that advantageous for any teacher? The more data you have, the better job you can do. It's like doing a research study that's incredibly biased towards one group, right? It's not really telling you anything important about the larger group. It's only telling you about that one biased group. So, um, that's what's missing from what we're doing right now."

Henry wanted to have access to his students' thinking because he intended to use what he learned to design in-class questions focused on what was challenging for students about the topic of epistasis. He explained that "in the future, then I can use that to make questions that target those ideas and bring them to the forefront a little bit better." He saw this as an effective approach because "until they start getting some questions wrong, they won't catch on that, 'Oh, maybe I don't understand this as well as I need to.'"

Ahead of his second semester teaching epistasis, Henry again reported that he was not sure what to expect regarding student thinking or how students would think about the epistasis question he would ask them to solve during class. This is reflected in the stagnation of his KST rubric score between semesters one and two (Figure 3.1). Henry said, "I imagine they've never talked about before the idea that one gene would completely interfere with another gene." During class, he planned to use an audience response system with multiple-choice questions to gather information from the whole class. Henry explained that the information elicited through a multiple-choice audience response system was useful but incomplete because "you know what they answered, but not why they answer." Therefore, he also planned to seek "person to person feedback" about whether students were confused by "floating around the room, hopping from person to person or group to group, just getting a feel for the class on what they're thinking, how they're setting a question up, things like that."

As he talked to students while they worked on an epistasis task in the second semester (Figure 3.2), Henry encountered an unanticipated pattern in students' problem-solving approach. Henry explained that the "big surprise" was that multiple students used a dihybrid cross to generate all possible genotypes from a two-locus cross, whereas, in the debrief, he had demonstrated a solution that used two monohybrid Punnett squares plus probability rules to arrive at the answer. He chose the monohybrid problem-solving approach because it seemed more straightforward to him and aligned with how he had taught his students to solve prior Mendelian tasks. However, after observing many students using a dihybrid cross, Henry explained that "in the case of epistasis, specifically, actually viewing them together does have, I think, a big advantage in that you can actually see the two genotypes next to each other because…one interferes with the other. It's more intuitive as to what's happening there rather than just purely going [with] math." Given this advantage, Henry was considering showing both approaches to solving the task in future semesters because "both [approaches] have pluses and minuses to them" and "it might be valuable for students to see that all of these [approaches] are accomplishing the same thing, just in different ways."

Ahead of the third semester that Henry taught epistasis, he anticipated how students would approach the topic, drawing on his prior experience. He said, "in past classes, [epistasis] is really kind of a brand new idea for most of them…the biggest issue that we encounter…was what's the best approach to the question?" He elaborated by saying "in any kind of genetic probability question, there's more than one approach you could take, and sometimes a certain approach is easier to understand than another approach…in the case of epistasis, it's probably easier to do them together. Or at least it's easier to visualize the outcome, if you do them together." This quote demonstrated that Henry retained the insights he had in the second semester about how students approached the epistasis task and he intended to act on the insight. Henry followed through on these plans, showing both approaches (single dihybrid cross and two separate monohybrid crosses) in the debrief portion after the original epistasis task in the third semester.

Henry also added a second task about epistasis in his third semester teaching the topic. Whereas the original question asked students to determine the probabilities of phenotypes for offspring based on parental phenotypes, the additional question offered a “realistic twist” by providing the phenotypes of offspring and asking students to derive the genotypes of the parents. In addition to feeling that the topic of epistasis needed more class time, Henry designed the question to push students beyond using the same “algorithmic” strategy for each problem. Henry explained that “I've noticed that when I give a question like this, where you throw a significant wrench into the algorithm machine, there's at least a good third of the class [that] really struggles with that.” He elaborated that “the wrench throwing is the only way to really, in my opinion, to really figure out the difference between a shallow algorithmic understanding of something and actually understanding the concept. If you just ask the same questions over and over again, just switching numbers and what organisms it's about and stuff, it's just applying the same algorithm over and over.” Here, Henry demonstrates an understanding of the importance of generative cognitive work, the kind of work that requires students to generate ideas and products that go beyond what has been presented to them, for developing deep conceptual understanding. Henry also demonstrated self-awareness of the limits of his knowledge of student thinking when he said, “I've not asked this question before, so I don't have any idea how it's going to go.”

Based on his observations during the target lesson in the third semester, Henry thought the new question had promise but needed to be “overhauled a little bit to make it a better question.” He observed that “the answer options were too easy... most of the students I talked to went based off of the answers that were available without really working through the problem.”

He described the thinking processes he had hoped for, but not achieved with the question like this, “So they don't really have to think about, ‘Okay, I have these two dogs. I don't know what their genotypes are. All I can do is observe the offspring. What does that tell me? Do I have some information, but not other [information]? What do I need to figure out?’ Instead they just look at the options and decide, ‘Well, that's the only one that can work.’ And so that's not the intent of that question…So this question would work much better as an open response. It would be difficult to write it as a truly good multiple-choice question. And I realized that after the first time I asked it.”

Lastly, Henry planned to create another epistasis task for the subsequent lesson a few days later, in order to provide students with more practice completing dihybrid crosses and to test their thinking about epistasis in a new context. Though he had observed that students did well on the epistasis tasks, he wondered whether they “really have a pretty good understanding of basic epistasis” or “this particular example is just very easy to understand.” If the latter were true, he expected that a question in a new context “wouldn't go quite as well.” Henry added that “there’s only one way” to find out “and that’s to try it.”

Synopsis: Teaching knowledge and practice associated with Henry’s PCK development

Henry built PCK by relying on coherent and interconnected knowledge and actions (Figure 3.3). Henry’s pedagogical knowledge of how people learn led him to embrace the idea that students learn by engaging in difficult tasks during class that require reasoning rather than regurgitation. In Henry’s thinking, difficult tasks created essential learning opportunities for students, and enabled both he and students to distinguish whether students had achieved the deep understanding of a topic that he desired. Based on this idea, he aimed to design tasks focused on the most challenging aspects of a topic. He considered the sequence of ideas and approaches students would need to employ in order to successfully complete a task. He then designed and refined tasks to ensure students engaged in the challenging thinking opportunities he intended to create for them, rather than circumventing the challenge. For Henry, involving students in generative work, such as tasks that require them to deal with a “wrench” thrown into their problem-solving algorithms, was paramount to helping students learn in his class. In order to create these challenging learning opportunities for students, Henry aimed to learn what was especially difficult about a topic for his students. In other words, he aimed to develop student-validated KST. He made guesses about what was difficult for students in the absence of other information and treated his guesses as hypotheses to be tested against empirical evidence of student thinking.

Henry used a two-pronged approach to gather this evidence of student thinking during class, which relied on his knowledge of monitoring student thinking. He collected some data from all students, using an audience-response system. This strategy followed from his concern that he may not hear a representative sample of student thinking if he interacted with only a subset of students. Yet he also saw limitations in this information, because it failed to reveal the logic and reasoning behind students’ answer choices. Henry wanted to identify the specific causes of and solutions for the difficulties students encountered with each topic. Therefore, in the second prong of his approach, he aimed to elicit and monitor student reasoning during class. He spent considerable time each class period interacting with students as they worked on tasks individually and in small groups. He talked to groups for several minutes, often asking multiple questions to probe their reasoning. Drawing on his pedagogical knowledge of how people learn, Henry’s primary goal in these interactions was to invite student reasoning, which he thought would facilitate deep student learning and help him construct his own KST. This could be contrasted with interacting with students in order to make sure they are on track or to replace students’ incorrect ideas with canonical ideas. Henry used what he learned from hearing student’s reasoning to inform his KIS (i.e., developed tailored KIS), which enabled immediate and future revisions to his teaching.

There are two overarching themes that emerge from how Henry developed PCK. First, implicit in how Henry taught and talked about his teaching was the idea that student thinking lies at the heart of teaching. In contrast, other participants seemed to place content at the heart of their teaching. Henry understood that difficult tasks help students learn and helped him learn about student thinking. Henry felt he was missing critical information when he did not know how students would think about a topic or approach a problem. Eliciting student thinking, especially their reasoning, was a top priority for Henry as he planned and implemented his target lesson. Second, Henry employed the mentality of a researcher in his teaching. He made and tested hypotheses about student thinking. He treated student thinking as data that was valuable to him and carefully considered how to minimize bias and increase the utility of the data he collected. He also treated his PCK as tentative, constantly aiming to gather more data about student thinking and revise what he knew. Henry developed PCK by intentionally creating and seizing opportunities to investigate students' thinking during challenging tasks, and then used evidence of their thinking to directly inform instructional strategies tailored to student thinking.

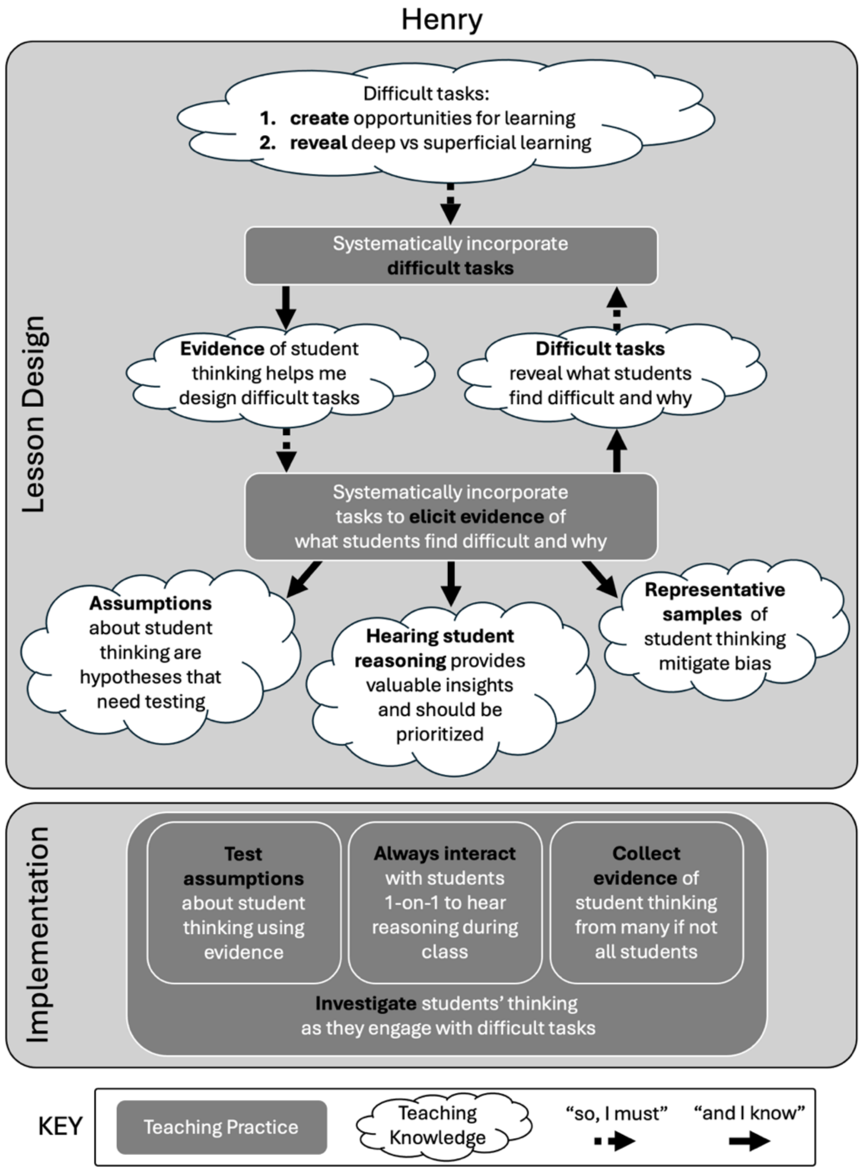


Figure 3.3. Visual representation of teaching knowledge and practices that supported PCK development for Henry. Both lesson design and implementation in the classroom contributed to PCK development. Teaching knowledge and practices that Henry relied on in lesson design then informed their practices during implementation. Each thought bubble (white) represents a component of a participant's pedagogical knowledge that informed his teaching practices. Rectangular boxes (dark gray) represent teaching practices undertaken in lesson design and implementation that contributed to PCK development. Solid and dashed arrows depict the relationships between teaching practices and supporting knowledge (see key). Bolded words improve readability and emphasize key components of knowledge and practice. For an interactive web-version of Figures 3 & 5 showing supporting qualitative evidence from each case study, visit https://awaugh19.github.io/interactive_PCK_case_figures.

PCK Development Case 2: June teaches molecular regulation

This case documents how June developed PCK for teaching the lac operon across two semesters of a microbiology course, starting with the first term she ever taught this topic and course.

Target lesson context: June's lesson information and teaching practices

Though new to the focal course, June had previously taught an introductory laboratory course focused on microbiology. The focal course enrolled roughly ~150 students per semester in 50-minute class periods. In the first semester studied, June abided by social distancing recommendations associated with the COVID-19 pandemic by remaining at the front of the classroom throughout class. This limited her access to student thinking.

In both semesters studied, June taught the lactose (lac) operon using an in-class activity within a larger unit about molecular regulation. The target lesson also included other gene regulatory systems, such as the lux operon and a short section about molecular regulation via small RNA (sRNA) molecules. In both semesters, June prepared students for a lac operon activity by explaining the mechanism of the lac operon using a diagram from their textbook as a visual aid. June explained how the lac operon enables *E. coli* cells to regulate lactose metabolism based on its availability and the absence of preferred nutrients like glucose. June explained, in detail, that the lac operon, in the presence of glucose, is repressed, inhibiting the cell's ability to consume lactose and resulting in cells only consuming glucose. When glucose is no longer available and the cells detect the presence of lactose, the lac operon is de-repressed, allowing the cell to begin consuming lactose. If colonies of *E. coli* are grown in a petri dish containing glucose and lactose, the colony will exhibit a characteristic two-phase growth curve in which cells grow for certain period until glucose is exhausted, then there is a short lag phase as the lac operon is de-repressed before a second growth period occurs, powered by lactose consumption.

After these detailed explanations, June engaged students in an activity about the lac operon. In the first semester, June's in-class activity involved a graph depicting the two-part growth curve. June verbally asked students, "Why is there a lag phase of growth after glucose is exhausted?" and showed the same question on the slide above the growth curve. June aimed for students to recognize that the lag phase is caused by the time it takes for de-repression to occur and for lactose-powered growth to begin. In the second semester, June kept the open-ended question and added three multiple choice questions that prompted students to describe which of three labeled points along the growth curve correspond to specific events in the activity of the lac operon (Figure 3.4).

Students in June's class spent much of class listening to her lecture and working either independently or in small groups on in-class activities. While students worked, June remained at the podium and occasionally circulated the room to talk with one or two groups of students about their work on the activity. Systematic analysis of video recordings from her lessons showed that students worked on in-class activities for 34% and 56% of two-minute segments during her first and second semesters, respectively, compared to an average of 25% across target lessons in our sample. June had one-on-one interactions in 4% and 20% of two-minute segments and heard students explain their thinking in 0% and 20% of two-minute segments during her first and second semesters, respectively, compared to the average of 8% across the dataset. These shifts likely relate to the lifting of social distancing restrictions, enabling her to leave the front of the room.

1. You can expect the CRP activator protein to be activated at:

A. Point A only
B. Point B only
C. Point C only
D. Points B and C only
E. Points A, B and C

2. You can expect the repressor protein to be produced at:

A. Point A only
B. Point B only
C. Point C only
D. Points B and C only
E. Points A, B and C

3. You can expect the repressor protein to be bound to the operator at:

A. Point A only
B. Point B only
C. Point C only
D. Points B and C only
E. Points A, B and C

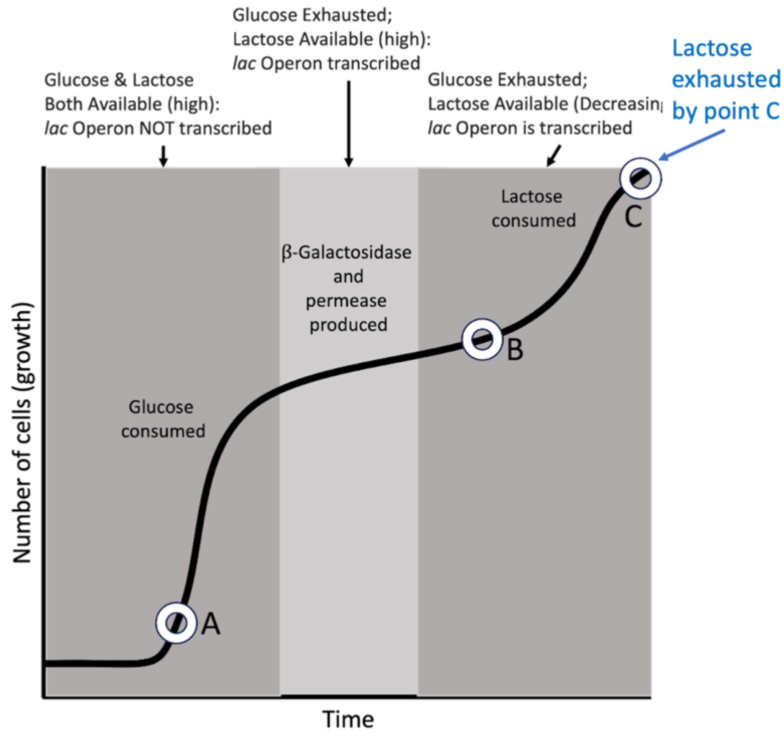


Figure 3.4: June's lac operon in-class activity used in the second semester studied. June used these questions in an exam in the first semester and then added these as in-class questions in the second semester. Students discussed the questions and shared answers via raising hands. Correct answers provided for the reader in green text but were not shown to students until debrief. In both semesters, June showed a less annotated version of this image in class and posed this question to students: "Why is there a lag phase of growth after glucose is exhausted?" In the second semester, June also added a new label and arrow (blue). Image and questions adapted by us from Campbell Biology (9th edition).

PCK development summary: The development for teaching the lac operon

In the first semester of data collection, June did not have extensive student-validated KST or integrated KIS about the lac operon, which is reflected in relatively low PCK scores (Figure 3.1, see semester 1). June relied on prior teaching experiences to anticipate student's thinking about the lac operon. She explained, "I've never taught this particular course before, but I've taught genetics…, and so usually [regulation] is where student's get the most confused... as far as regulation goes, if it is touched on in previous courses, it's like a very small like, 'Hey, this thing exists, but we're not going to go into the details of how it works.'" Based on her anticipation that students would be unfamiliar with the lac operon, June planned to "slow down" during this portion of the lesson and "include at least one active-learning strategy." She hoped that incorporating formative assessment to teach this topic would provide structure for "immediate feedback from [students]" to "in the moment, see where they're at" and would help her assess whether she was "doing a good job explaining it." June explained that these opportunities for formative assessment also improved students' attention and engagement with the material, "it helps them stay more engaged in that particular topic if they know that they're like having to answer a question."

June wanted to know how students arrived at answers to questions she posed in class but lamented that "in a class of 200[1] students, it's hard to hear what their logic is. I just have to kind of guess." After inviting students to answer a question in small groups, June typically invited a volunteer to answer the question for the class, providing a brief glimpse into one student's thinking. June worried that "when I'm just getting verbal feedback, it's only the people that are confident in their answer that they're speaking up. So, I don't know that that's the best snapshot." To get a better snapshot of the class, June occasionally asked students to raise their hands to indicate an answer. June did not interact with students as they worked on questions in class in the first semester but found opportunities to hear student reasoning in other contexts, "it's more so at the office hours that I get to peek into individual student logic behind how they're arriving at certain questions."

Ahead of the second semester, June gained new insights into the challenges students faced learning about the lac operon, based on evidence from their poor performance on a set of exam questions and a few conversations during office hours. After reflecting on this evidence and her prior lesson, June reasoned that, "[students] have difficulty when we reach those higher

---

[1] June later reported her class enrollment as 150, as noted in Table 4.2

levels of combining concepts together. And so rather than expecting them to just put the individual pieces together and apply it on an exam question, which is kind of what I was expecting them to be able to do last [time I taught the course], [I had] this realization that no, they can't just jump to that. I didn't really prepare them for that." In response, June decided to use the three difficult exam questions as an in-class activity in order to "bridge the gap" so that students would be prepared for questions that required "application or synthesis." In addition to using the exam questions as an in-class activity, June modified the graph associated with the questions based on conversations she had with students in office hours (Figure 3.4). She had learned that students were confused about whether lactose was still present at a particular point because the growth curve was not perfectly flat at the end of the experiment. June recalled that a student in office hours thought, "Well, it's not necessarily in stationary phase yet. It's not clear that by the time that time block is over, that lactose is completely gone because it says it's available and it doesn't necessarily say that it's all gone." June readily recalled her response to the student, "And I'm like, 'Okay. Yeah, you're totally right. That's an acceptable answer based on your interpretation of the graph. Got it.'" To address this issue, June added an arrow with accompanying text to denote the time point of the growth curve at which lactose was exhausted (Figure 3.4).

June sought more access to student thinking during class in the second semester, but also experienced a tension about eliciting students' reasoning. June interacted with three different student groups, a change from the first semester's target lesson in which, due to social distancing guidelines, June had not interacted with students working in small groups. June explained that she could learn about students' misconceptions from hearing their reasoning in small groups, "if they defend [their answer], especially if it's a wrong answer, if they defend it, it can tell me at what point the misconception is forming." However, June also felt that she was inconsistent in prompting students to share their reasoning and instead sometimes prioritized her own explanations, "During that clip, in contrast to the earlier clip, I did most of the talking and the explaining after they did the easier side of things [...] I felt like I was patient, I left the door open for them to give an answer, but when the answer wasn't necessarily what I was expecting, I kind of scrambled to be like, ‘Okay, well, let me just tell you the answer sort of thing.’” We asked June why this stood out to her, and she responded, “Because it's not really forcing their brains to do the work. They're just passively listening to my explanation, and it doesn't really give me a peek into their thought process.” This demonstrates that June thought she *and* students could benefit from students sharing the logic behind their answers during class, and that she sometimes struggled to prioritize this intention.

June also expressed internally inconsistent ideas about how she could best interact with small groups. Though quotes above show that she valued students explaining their reasoning, she described her ideal interaction with students this way, “If I had all the time in the world, I would baby step them and say, ‘Okay, what is that part of the image representing?’ and ‘Okay, as a result, what's it doing here?’ And like the leading questions, where, rather than just saying, ‘Dive into it and you figure it out’. Like walking them step-by-step of like, ‘Okay, let's focus over here first. Let's make sure you understand that before we move on to how that impacts whether or not it's transcribed.’” In this hypothetical example, June would ask many questions to extract a full explanation from students, which has the potential to undermine her intentions to make students’ brains do the work and to elicit insights about student reasoning.

Moving forward, June planned to keep the former exam questions as an in-class activity and cut some content. She explained that it was “eye-opening to see this wasn't just a few students misinterpreting this graph last year. They're all taking a significant amount of time to really understand what the graph is representing before we can actually dive into the questions. So, I felt like that was time well spent.” She also planned to cut some content but was still making those decisions. She explained, “Something’s got to go, but it’s a matter of what. Some of these learning objectives can actually be achieved in their homework, and I don't necessarily have to do it in lecture.” June planned to make these cuts to the content that needs “less building” so that she could “use the time in lecture where I can actually chatter with them for the harder topics.”

Synopsis: Teaching knowledge and practice associated with June’s PCK development

June built PCK by taking advantage of unexpected encounters with student thinking and translating what she learned into lesson revisions (Figure 3.5). She displayed both curiosity and appreciation for student thinking when she described her intentions for in-class interactions with small groups and attended to evidence of student thinking in exam performance and office hours. June immediately used new student-validated KST to modify her instructional strategies for teaching the lac operon (i.e., developed tailored KIS). She also interpreted the time that students needed to complete the new in-class questions, which had been challenging for students on the previous semester’s exam, as evidence that the learning task was worthwhile. This suggested that she has budding intentions to provide students with practice engaging in challenging tasks.

June’s pedagogical knowledge of monitoring student thinking allowed her to capitalize on opportunities to learn about student thinking when faced with them, but she did not proactively seek student reasoning when it was most available to her – during class. In the first semester, June heard students’ ideas from the front of the room and these answers were rarely more than a handful of words. In the second semester, June interacted more often with students, but often undermined the value of these interactions for her own learning by doing more talking than listening. June leaned heavily on students’ volunteered questions and answers shared during debrief portions of her in-class tasks. She felt limited by her class size, which was very similar to Henry’s, and worried about bias in volunteered responses, but had not yet implemented strategies to gather a more representative sample of student thinking, such as an audience response system.

Overall, June's approaches to collecting evidence of student thinking during class suggested an intention to get a snapshot of student thinking in order to evaluate the effectiveness of her own teaching, rather than an intention to learn the details of students' reasoning about the topic. One result of this is that her tailored KIS rested on the knowledge that students had struggled with particular questions on an exam, but she lacked insight about why students found the questions difficult or what sort of thinking they engaged in to answer the questions.

June was able to develop PCK over her first two semesters because she was curious about student thinking, saw value in it for her instruction, and felt drawn to incorporate a challenging set of questions in her lesson. Despite June's impressive PCK development between her first two semesters teaching the lac operon, her capacity for future PCK development may be undermined by particular practices and ways of thinking. June held internally inconsistent ideas that limited her access to student thinking. She felt that she needed to explain concepts, while also demonstrating knowledge of monitoring student thinking that motivated her to hear students' explanations. In practice, these internally inconsistent ideas often compromised June's opportunities to get access to students' reasoning in class. June's PCK development may also be limited by her focus on in-class tasks as useful for engaging students' attention, rather than as essential opportunities to engage students in generative work focused on the most difficult aspects of a topic. Yet June also displayed nascent pedagogical knowledge of how people learn that included an awareness that students' brains needed to do the work, which could lead her to continue to prioritize in-class tasks that prove challenging for students.

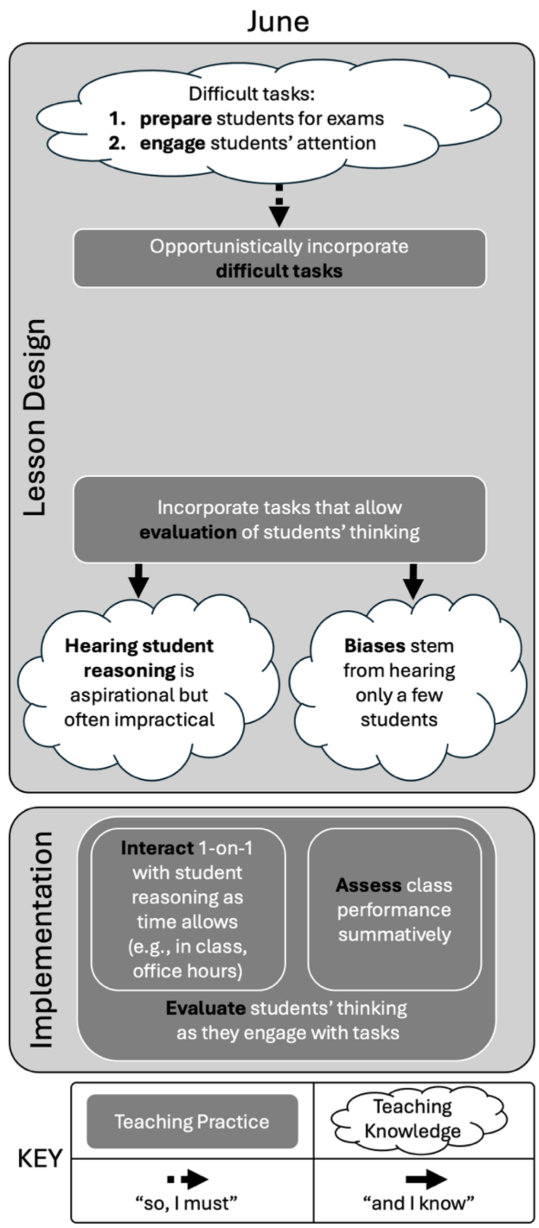


Figure 3.5. Visual representation of teaching knowledge and practices that supported PCK development for June. Both lesson design and implementation in the classroom contributed to PCK development. Teaching knowledge and practices that June relied on in lesson design then informed their practices during implementation. Each thought bubble (white) represents a component of a participant's pedagogical knowledge that informed his teaching practices. Rectangular boxes (dark gray) represent teaching practices undertaken in lesson design and implementation that contributed to PCK development. Solid and dashed arrows depict the relationships between teaching practices and supporting knowledge (see key). Bolded words improve readability and emphasize key components of teaching knowledge and practice. For an interactive web-version of Figures 3 & 5 showing supporting qualitative evidence from each case study, visit https://awaugh19.github.io/interactive_PCK_case_figures/.

Discussion

This longitudinal investigation of PCK development is the first among undergraduate STEM instructors and the largest study of PCK development to date at any educational level. We observed variation in PCK that warrants attention in future research, including whether an instructor's knowledge of student thinking (KST) is grounded in evidence from students (i.e., student validated) and the extent to which their knowledge of instructional strategies (KIS) draws on KST (i.e., integrated) to enact strategies tailored to student thinking. The findings also highlight that while participants tended to develop PCK that was more student validated and integrated, teaching experience did not consistently foster knowledge development. Some instructors built PCK more consistently and efficiently than others, and the case studies point to factors influencing PCK development. In this section, we situate our findings in the literature and explore research directions enabled by the qualitative features of PCK identified in this study—student-validated, broad-strokes, and tailored PCK. Then, we discuss ways that pedagogical knowledge influenced PCK development, as revealed by our case studies, before presenting a hypothetical model of how knowledge and teaching practices support PCK development. We conclude with practical advice from this research aimed at instructors seeking to develop their PCK and those who design and facilitate teaching professional development.

Research implications of PCK features: student validated, broad strokes, and tailored

Prior studies of PCK have not focused on the source of information instructors rely on for their KST (e.g., Andrews et al., 2022; Chan & Hume, 2019). It is not clear if the PCK elicited from instructors in prior PCK scholarship was consistently student validated or if the source was not a point of inquiry. Looking closely at quotes demonstrating KST in published studies, we observed this knowledge was often student validated (e.g., Chan & Yung, 2018a; Park & Chen, 2012; Park & Oliver, 2008). It may be the case that K12 educators are positioned to quickly develop student-validated PCK, due to their formal training as teachers and teaching contexts that provide closer teacher-student interactions. This could explain why prior research in K12 contexts has not reported variation in the source of KST and point toward an important difference in PCK between undergraduate and K12 teachers. It may also be true that the source of KST varies among K12 teachers and warrants further investigation.

We propose that future PCK research, regardless of educational level, specifically consider whether KST is student validated, and the impact of the source of KST on teaching practices and student outcomes. Two hypothesized benefits of student-validated KST underscore the need for future work in this area. First, misconception-focused instruction can facilitate learning gains that go above and beyond the benefits of active learning (Nehm et al., 2022) and instructors would likely need student-validated KST to design this kind of instruction. Second, we posit that developing student-validated KST forms a reliable and actionable basis on which to design instruction, when compared to assumptions and guesses about student thinking. Student-validation confers reliability by providing evidence that some students relied on a particular way of thinking or set of ideas during the lesson. Student-validation confers actionability because it provides feedback about how students responded to a specific instructional strategy, enabling an instructor to modify that strategy to better support students in achieving the learning objectives. Instructors may need support noticing student thinking in a way that can help them build student-validated KST. For instance, pre-service K12 teachers identified student thinking evidence from non-verbal cues more readily than from students' verbal or written work (Lam & Chan, 2020), yet non-verbal cues are likely to provide less reliable information. about student thinking.

PCK research methodology may particularly stand to benefit from distinguishing student-validated KST in circumstances where the accuracy or comprehensiveness of the instructor's KST cannot be assessed. Some PCK studies assess instructors' PCK accuracy by comparing it to the findings of systematic investigations of students' thinking (Park et al., 2018). In our study, instructors taught a wide range of content, most of which has not been the subject of empirical research about student thinking. Thus, the available canonical PCK provided insufficient evidence to characterize the accuracy and comprehensiveness of participants' KST. Since student-validated KST offers a qualitative indicator that the instructor observed a certain student thinking pattern, student-validated KST could serve as a useful proxy for assessing the accuracy of an instructor's KST in the many cases where no systematic data about student thinking exists as a comparator.

In this study, instructors exhibited KIS integrated with KST as well as KIS lacking integration, which is similar to previous results in K12 educational contexts (e.g., Park & Chen, 2012; Sickel & Friedrichsen, 2018). However, in identifying the distinction between broad-strokes and tailored KIS, we uncovered previously undescribed variation in PCK integration. Looking closely at quotes demonstrating KIS-KST integration in prior work, we observed that they often involved tailored integration (e.g., Park & Chen, 2012; Park & Oliver, 2008). It is unclear if K12 teachers largely rely on tailored KIS, resulting in prior research that does not make this distinction, or if K12 teachers also use KIS with different levels of specificity. The rubric that we developed builds upon the PCK mapping approach, which counts instances of integration (e.g., Park & Chen, 2012), by distinguishing between broad-strokes and tailored integration and could be used to test for this variation in future research among K12 and undergraduate teachers.

Importantly, it is not clear that tailored KIS necessarily leads to more effective teaching than broad-strokes KIS. This too should be investigated in future research. For example, some naive ideas that students express in class may be relatively easy to help them reconsider. In those cases, a general strategy like repeating an idea or giving more examples, could efficiently and effectively support students' knowledge construction. However, we propose that PCK among highly effective instructors would not exclusively consist of broad-strokes KIS because some naive ideas and difficulties that students' experience require carefully designed instruction that targets the details of student thinking (e.g., Kalinowski et al., 2010; Kalinowski et al., 2013; Nehm et al., 2022). Additionally, higher-quality instructional moves among pre-service teachers have their roots in more detailed interpretations of student thinking (Lam & Chan, 2020). The field would benefit from future research that explored the impact of broad-strokes and tailored KIS on teaching practices and student outcomes, including a focus on the contexts in which each kind of knowledge proves most beneficial to student learning.

Pedagogical knowledge and associated teaching practices facilitate PCK development

Our case studies indicate that pedagogical knowledge can have considerable influence on PCK development. The vast majority of prior research about PCK among K12 and undergraduate instructors has not considered a role for pedagogical knowledge. Additionally, compared to PCK scholarship, many fewer studies of teacher expertise consider pedagogical knowledge (Andrews et al., 2022). Our cases of two instructors join a growing body of work that finds a key role for pedagogical knowledge in PCK development (e.g., Andrews et al., 2019; Chan & Yung, 2018a; Gehrtz et al., 2022). In particular, our cases echo and expand upon scholarship that suggests efficient PCK development follows from accessing and interpreting student reasoning as students engage in challenging questions (e.g., Andrews et al., 2019; Chan & Yung, 2018a; Copur-Gencturk & Atabas, 2024). The fact that multiple studies have revealed similar patterns of pedagogical knowledge and teaching practices that facilitate PCK development is significant because instructors with greater PCK can foster greater student learning (e.g., Hill et al., 2005; Kanter & Konstantopoulos, 2010; Sadler et al., 2013).

Multiple lines of evidence suggest a role for pedagogical knowledge of monitoring student thinking in PCK development (Andrews et al., 2019; Chan & Yung, 2018a, 2018b; Gehrtz et al., 2022). In our case studies, we add to this work by identifying specific ways in which knowledge of monitoring student thinking supported teaching practices that facilitated PCK development (Figures 3 & 5). Both Henry and June were sometimes surprised by what they learned when they considered evidence of student thinking. Henry did not anticipate the approaches that students would use to solve a problem, and June did not anticipate how difficult questions would be, both on the exam and later when she brought them into class. This emphasizes that instructors are unlikely to be able to anticipate the complexity in how students think about a topic or question, let alone the variation among students, especially when they are new to teaching a topic. Furthermore, placing student thinking as a top priority led to different instructional strategies and different opportunities to develop PCK in our cases. Thus, we echo the call for instructors, educational developers, *and* researchers to prioritize student-thinking-centered instruction, not just student-centered instruction (Gehrtz et al., 2022). Student-thinking-centered instruction involves instructors prioritizing student thinking in their teaching by intentionally accessing student thinking, working to interpret students' reasoning (rather than just evaluating accuracy), and responding to what they learn about student thinking.

Pedagogical knowledge of how people learn and similar constructs, such as constructivist orientations or beliefs, also seem to serve an important role in PCK development (Andrews et al., 2019; Auerbach & Andrews, 2018; Brown et al., 2013; Friedrichsen et al., 2011; Magnusson et al., 1999; Sickel & Friedrichsen, 2018). Instructors learned what challenged students about a topic when they asked questions that required generative work (i.e., generating something beyond what was presented to them), and they learned how students reasoned through those challenges when they asked students to externalize their reasoning (Figures 3 & 5). Questions that students answered easily failed to illuminate the bounds of students' thinking and questions that did not elicit reasoning left the instructor with few insights. Therefore, asking students to engage in generative questions and externalize their reasoning created conditions for developing student-validated KST. This is an exciting finding because knowledge of how people learn has the potential to both increase student learning through more effective teaching practices (e.g., Chi & Wylie, 2014) and to facilitate instructor's PCK development, which can also increase student learning. Future work should continue to unravel the relationship between general pedagogical knowledge of how people learn and KIS, including whether an instructor's ability to implement knowledge of how people learn is modulated by their

PCK for a given topic. For instance, a teacher's ability to design generative tasks may be aided by their PCK, as PCK encompasses topic-specific knowledge of which areas of the content students find difficult and how to support their learning in those areas.

Hypothetical model of pedagogical knowledge, PCK, and PCK development

We propose a hypothetical model of how pedagogical knowledge and resulting teaching practices facilitate PCK development, and how PCK itself supports further PCK development. (Figure 3.6). This model draws on findings from this study and prior work at both the K-12 and undergraduate level (Andrews et al., 2019; Auerbach & Andrews, 2018; Chan & Yung, 2018a, 2018b; Gehrtz et al., 2022). In the model we propose, pedagogical knowledge and PCK inform instructional strategy design to elicit evidence of student thinking about a topic, with an emphasis on what students find especially difficult. Furthermore, PCK aids instructors as they interpret the evidence of student thinking they access during class (e.g., Johnson & Larsen, 2012; Speer & Wagner, 2009; Wagner et al., 2007). In turn, this interpretation informs future instructional design and implementation.

In the planning phase, an instructor draws on their pedagogical knowledge of how people learn and monitoring student thinking to design instructional strategies that create opportunities for students to engage in generative work and position the instructor to encounter student reasoning. Planning of instructional strategies also involves PCK as instructors can deploy their KIS and KST to design learning opportunities focused on what is most challenging for students about a particular topic.

The model also hypothesizes that instructors draw on their pedagogical knowledge and PCK as they facilitate students' work on planned instructional strategies during class (Figure 3.6). Using the knowledge that students learn by constructing their own ideas, instructors aim to maintain the onus on students to do the generative work of articulating their reasoning, rather than providing hints or questions that guide a student directly to particular ways of reasoning (e.g., Tekkumru-Kisa et al., 2022). Using knowledge of monitoring student thinking, instructors pay attention to and value the idiosyncrasies and patterns of student thinking that emerge when they invite multiple students to share their reasoning (e.g., Gehrtz et al., 2022). PCK also informs the implementation of planned instructional strategies as it enables instructors to anticipate and interpret student reasoning in relation to anticipated topic-specific difficulties.

A key outcome of this design and implementation, beyond valuable opportunities for student learning, is access to evidence of student thinking for the instructor. Here again, PCK aids instructors by enabling more efficient interpretation of student thinking, even early and ill-formed ideas (e.g., Andrews et al., 2022). PCK can influence which aspects of student thinking instructors prioritize for follow-up, whether they recognize a common misconception embedded in student reasoning, and what they can conclude from students' performance on an in-class multiple choice question. The PCK that instructors develop from evidence of student thinking can then inform planning as instructors change or build on a designed instructional strategy. Thus, PCK itself can facilitate further PCK development as instructors leverage their KST and KIS to more efficiently access, interpret, and respond to the student reasoning they elicit (e.g., Gehrtz et al., 2022; Johnson & Larsen, 2012; Speer & Wagner, 2009; Wagner et al., 2007). Future work should consider testing this hypothetical model as we continue to unravel the relationships between pedagogical knowledge, PCK, and PCK development.

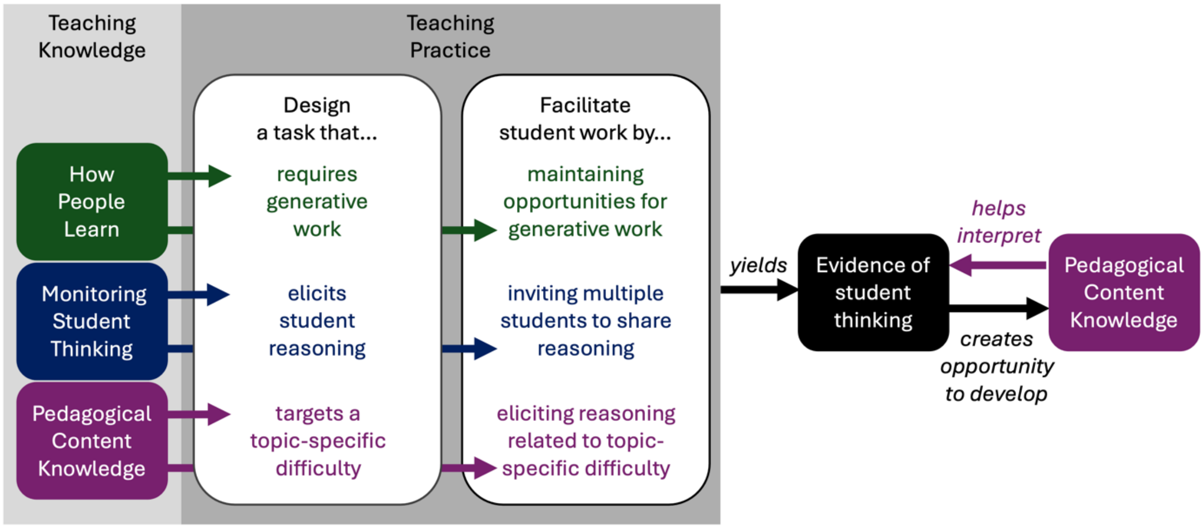


Figure 3.6. Hypothetical model depicting the role of knowledge and teaching practices in PCK development. In concert with PCK (purple), pedagogical knowledge of how people learn (green) and of monitoring student thinking (blue) inform practices as instructors design in-class tasks and facilitate students' work on those tasks. These efforts yield evidence of student thinking (black) which creates an opportunity for the instructor to develop more PCK. Instructor's PCK also helps them interpret evidence of student thinking in the process of PCK development. Unless otherwise labeled, arrows connect teacher knowledge to their hypothesized translation into instructional design and implementation.

Implications for undergraduate instructors and those who support instructor development

As previous sections outline, there are many avenues for future research to test and elaborate on the findings of this work, but instructors may be interested in reflecting on their own teaching more immediately. In that spirit, we offer a set of reflection questions and actions, inspired by our participants, that instructors could employ and that could be useful in teaching development programming for future and current college instructors (Table 3.5). One source of motivation for these questions comes from the interview protocols used in this study (Appendices A & B). Some participants noted that engaging in interviews for this study fostered their own reflection on their teaching, and Table 3.5 makes these same opportunities available to readers, along with ideas for immediate actions.

This research examined college biology instructors as learners, and we encourage instructors, educational developers, and researchers to similarly recognize that teaching requires constant learning. When we view instructors, including ourselves, as learners, we can benefit from the insights of vast scholarship about how people learn and expertise development. Ongoing expertise development is often marked by coexistence of and competition between productive and unproductive ideas within the learner's mind (Nehm & Ridgway, 2011), which is a possible driver of misalignment between thinking and practice. Interestingly, June recognized just such a tension between her own teaching knowledge and practices when she defaulted to providing an explanation, rather than keeping the onus on the student to provide their reasoning. June acknowledged that this meant that the student was “just passively listening to [her] explanation,” meaning she could not get “a peek into their thought process.” Crucial moments for expertise development can happen when, like June, learners recognize these tensions themselves. Professional development can support this kind of self-motivated change by emphasizing that collecting information about student thinking is not just an advanced practice, but a foundational practice that can enrich teaching from the outset. Early-career instructors can be empowered by understanding that, like the participants in this study, the path to leading effective classes, growing comfortable with teaching, and effectively facilitating student learning lies in actively engaging with student thinking.

Table 3.5. Reflection questions and possible actions for instructors to support PCK development.

Reflection questions	Possible actions
What do I know about how my students think about [topic]? What evidence or information convinces me that this is how my students	**Plan**: For one in-class task you plan to use, take five minutes to brainstorm the range of ideas and approaches students might use to complete the task. Then consider if you can revise the task to better elucidate student thinking or to test your hypotheses about student thinking.
	Teach: For an in-class task in your next lesson, set a goal to identify at least two student ideas that emerge as they work on the task.

think about [topic]?	**Reflect**: Spend five minutes after class writing down the patterns of student thinking you observed during class. Consider what tweaks you want to make to the next lesson (or this lesson next year) based on those observations.
What access do I have to student thinking? How could I expand that access?	**Plan**: Plan to walk around the class before the debrief portion and try to hear at least 2 students explaining their thinking during small group work.
	Teach: Ask students in small group or one-on-one settings (e.g., office hours, study sessions, in class) to explain their reasoning after they offer an answer. Focus on understanding rather than evaluating.
	Teach: Instead of hearing questions only from those willing to raise their hand, end class by asking each student to write down one question they have on a notecard and collect them.
	Reflect: Take notes on which circumstances and actions allowed you the most useful to access student reasoning during class.
Are there ways in which I'm unintentionally limiting my own access to evidence of student thinking? Are there instances where I provide my own reasoning instead of hearing from students?	**Plan**: Delineate which hints/explanations you are willing to provide to alleviate unproductive difficulty and which information you intend to withhold for the purpose of learning. Consider what questions you could ask, without being too leading, to help students work through difficulties.
	Teach: Prioritize one-on-one student interactions during class to maximize opportunities to access and understand student thinking. Resist the urge to supply information before hearing student reasoning. Instead, ask them to share how they are thinking about the question.
	Reflect: Review the circumstances and actions that led you to provide your own reasoning at the expense of student learning and compare it to the circumstances where you successfully elicited student reasoning.
Which of my instructional strategies integrate my knowledge of student thinking? Do they address which topics students find difficult and *why* they find them difficult?	**Plan**: Consider which patterns of student thinking are most important to address or leverage in relation to learning objectives. Design, select, or refine instructional strategies to target root causes of these patterns.
	Teach: Use formative assessment and other student feedback to inform in-the-moment instructional decisions. Address student thinking patterns with targeted examples and practice problems, rather than more of the same.
	Reflect: Review which instructional strategies are based on what you know about student thinking about a topic, and how specific they are to the details of students' thinking. Consider whether students would benefit from a strategy tailored to the specific difficulties they encounter.

Limitations

We caution the reader from interpreting results from this study without first considering its limitations. We did not set out to judge the accuracy of the participants' PCK, and instead aimed to characterize qualitative variation in PCK and how it changed over time. Therefore, some of the PCK we documented, even student-validated KST, may not align with what we would discover about student thinking in a systematic investigation. It was not possible to judge the accuracy and comprehensiveness of PCK because the undergraduate biology community has not agreed upon standard learning objectives for given courses or levels, nor have researchers systematically investigated student thinking about most biology topics.

Another potential limitation is the impact of the COVID-19 pandemic on longitudinal data collection. Most instructors had to adapt their target lesson to fit a new modality, which could have impacted their ability to interact with student thinking. Additionally, instructors may have been less attentive to student thinking during class because of the stressors associated with pandemic. We carefully considered how changes in instructional modality influenced our ability to elicit PCK and could not detect patterns that would invalidate our findings (Supplementary Figure S3.1).

Our sample of participants differs from the typical undergraduate instructor population in important ways, and therefore findings should not be generalized to all biology faculty. Many participants were engaged in teaching professional development, with most having over 40 hours of teaching professional development and a few who had experience with education research. This level of involvement suggests these instructors had a strong investment in improving their teaching and may have knowledge and skills that facilitate PCK development that not all faculty have yet had opportunities to develop. As a result, this study may overestimate the PCK more efficiently than is typical.

Since PCK is a topic-specific category of teacher knowledge, the PCK each participant exhibited in this study represents just a fraction of their overall PCK. We only collected data in relation to a single focal lesson, and the breadth and depth of participants' PCK could differ across the topics they teach. It is possible that they would have developed PCK differently if we focused on a different topic. We also note that some instructors participated in the study longer than others, giving certain participants more opportunities to exhibit PCK development.

Lastly, PCK researchers should keep in mind that we focused only on two components of PCK (KST and KIS), whereas the complete theoretical framework of PCK also includes knowledge of assessment and knowledge of curriculum (e.g., Gess-Newsome, 2015). We focused on KST and KIS because these components of PCK are the most studied and considered to be key components of PCK critical in shaping of an instructor's overall PCK (Andrews et al., 2022; Chan & Yung, 2015; Depaepe et al., 2013; Park & Chen, 2012; Park & Oliver, 2008). Our definition of KIS may overlap with knowledge of assessment because, in our participants' active learning classrooms, formative assessment often served as the instructional strategy. Nonetheless, a study that considered the four components of PCK distinctly might reveal additional or different findings about how instructors develop PCK.

Acknowledgements

We extend our gratitude to our research participants for their valuable time and energy. They stuck with the study over repeated semesters and through a global pandemic, and we feel honored to have had the chance to learn alongside them. We are also grateful to the University of Georgia's Biology Education Research Group (BERG) and the Andrews lab for their feedback and support. Dr. Cristine Donham provided important assistance with PCK rubric scoring. Dr. Sandhya Krishnan trained an initial undergraduate team on our observation protocol. Aman Patel led his fellow undergrads in implementing the protocol, including Esha Mohnalkar, Michella Obialor, Katia Uriostegui-Santos, Bryn Robinson. Thanks also to advisors Dr. Soonhye Park and Dr. Alicia Alonzo, who provided feedback on research methodology, and Dr. Alexandra Cooper, who provided feedback on the manuscript. Support for this work provided by the National Science Foundation's Improving Undergraduate STEM (IUSE) & ECR program under award No. 1845886. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Alonzo, A. C., Berry, A., & Nilsson, P. (2019). Unpacking the complexity of science teachers' PCK in action: Enacted and personal PCK. *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science*, 273-288.

Alonzo, A. C., & Kim, J. (2016). Declarative and dynamic pedagogical content knowledge as elicited through two video-based interview methods. *Journal of Research in Science Teaching*, *53*(8), 1259-1286.

American Association for the Advancement of Science (AAAS). (2011). *Vision and change in undergraduate biology education: A call to action*.

Andrews, T. C., Auerbach, A. J. J., & Grant, E. F. (2019). Exploring the relationship between teacher knowledge and active-learning implementation in large college biology courses. *CBE—Life Sciences Education*, *18*(4), ar48.

Andrews, T. C., & Lemons, P. P. (2015). It's personal: Biology instructors prioritize personal evidence over empirical evidence in teaching decisions. *CBE—Life Sciences Education*, *14*(1), ar7.

Andrews, T. C., Speer, N. M., & Shultz, G. V. (2022). Building bridges: a review and synthesis of research on teaching knowledge for undergraduate instruction in science, engineering, and mathematics. *International Journal of STEM Education*, *9*(1), 66.

Anfara, V. A., Brown, K. M., & Mangione, T. L. (2002). Qualitative analysis on stage: Making the research process more public. *Educational Researcher*, *31*(7), 28-38.

Auerbach, A. J. J., & Andrews, T. C. (2018). Pedagogical knowledge for active-learning instruction in large undergraduate biology courses: a large-scale qualitative investigation of instructor thinking. *International Journal of STEM Education*, *5*, 1-25.

Brown, P., Friedrichsen, P., & Abell, S. (2013). The Development of Prospective Secondary Biology Teachers PCK. *Journal of Science Teacher Education*, *24*(1), 133-155.

Carlson, J., Daehler, K. R., Alonzo, A. C., Barendsen, E., Berry, A., Borowski, A., Carpendale, J., Kam Ho Chan, K., Cooper, R., & Friedrichsen, P. (2019). The refined consensus model of pedagogical content knowledge in science education. *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science*, 77-94.

Chan, K. K. H., & Hume, A. (2019). Towards a consensus model: Literature review of how science teachers' pedagogical content knowledge is investigated in empirical studies. *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science*, 3-76.

Chan, K. K. H., & Yung, B. H. W. (2015). On-site pedagogical content knowledge development. *International journal of science education*, *37*(8), 1246-1278.

Chan, K. K. H., & Yung, B. H. W. (2018a). Developing pedagogical content knowledge for teaching a new topic: More than teaching experience and subject matter knowledge. *Research in science Education*, *48*, 233-265.

Chan, K. K. H., & Yung, B. H. W. (2018b). Pedagogical Content Knowledge Development in Experienced Biology Teachers in Their First Attempts at Teaching a New Topic. In J. Yeo, T. W. Teo, & K.-S. Tang (Eds.), *Science Education Research and Practice in Asia-Pacific and Beyond* (pp. 197-214). Springer Singapore.

Chi, M. T., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational psychologist*, *49*(4), 219-243.

Clemmons, A. W., Timbrook, J., Herron, J. C., & Crowe, A. J. (2020). BioSkills Guide: Development and National Validation of a Tool for Interpreting the Vision and Change Core Competencies. *CBE—Life Sciences Education*, *19*(4), ar53.

Copur-Gencturk, Y., & Atabas, S. (2024). A microgenetic analysis of teachers' learning through teaching. *International Journal of STEM Education*, *11*(1), 29.

Dancy, M., Henderson, C., & Turpen, C. (2016). How faculty learn about and implement research-based instructional strategies: The case of peer instruction. *Physical Review Physics Education Research*, *12*(1), 010110.

Depaepe, F., Verschaffel, L., & Kelchtermans, G. (2013). Pedagogical content knowledge: A systematic review of the way in which the concept has pervaded mathematics educational research. *Teaching and teacher education*, *34*, 12-25.

Friedrichsen, P., Driel, J. H. V., & Abell, S. K. (2011). Taking a closer look at science teaching orientations. *Science education*, *95*(2), 358-37

Gehrtz, J., Brantner, M., & Andrews, T. C. (2022). How are undergraduate STEM instructors leveraging student thinking? *International Journal of STEM Education*, *9*(1), 18.

Gess-Newsome, J. (2015). A model of teacher professional knowledge and skill including PCK: Results of the thinking from the PCK Summit. In *Re-examining pedagogical content knowledge in science education* (pp. 28-42). Routledge.

Gregory, T. R. (2009). Understanding natural selection: essential concepts and common misconceptions. *Evolution: Education and outreach*, *2*, 156-175.

Guba, E. G. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries. *Ectj*, *29*(2), 75-91.

Henze, I., van Driel, J. H., & Verloop, N. (2008). Development of Experienced Science Teachers' Pedagogical Content Knowledge of Models of the Solar System and the Universe. *International journal of science education*, *30*(10), 1321-1342.

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American educational research journal*, *42*(2), 371-406.

Johnson, E. M., & Larsen, S. P. (2012). Teacher listening: The role of knowledge of content and students. *The Journal of Mathematical Behavior*, *31*(1), 117-129.

Kalinowski, S. T., Leonard, M. J., & Andrews, T. M. (2010). Nothing in evolution makes sense except in the light of DNA. *CBE—Life Sciences Education*, *9*(2), 87-97.

Kalinowski, S. T., Leonard, M. J., Andrews, T. M., & Litt, A. R. (2013). Six classroom exercises to teach natural selection to undergraduate biology students. *CBE—Life Sciences Education*, *12*(3), 483-493.

Kanter, D. E., & Konstantopoulos, S. (2010). The impact of a project-based science curriculum on minority student achievement, attitudes, and careers: The effects of teacher content and pedagogical content knowledge and inquiry-based practices. *Science education*, *94*(5), 855-887.

Lam, D. S. H., & Chan, K. K. H. (2020). Characterising pre-service secondary science teachers' noticing of different forms of evidence of student thinking. *International journal of science education*, *42*(4), 576-597.

Magnusson, S., Krajcik, J., & Borko, H. (1999). Nature, sources, and development of pedagogical content knowledge for science teaching. In *Examining pedagogical content knowledge: The construct and its implications for science education* (pp. 95-132). Springer.

Nehm, R. H., Finch, S. J., & Sbeglia, G. C. (2022). Is active learning enough? The contributions of misconception-focused instruction and active-learning dosage on student learning of evolution. *BioScience*, *72*(11), 1105-1117.

Nehm, R. H., & Reilly, L. (2007). Biology majors' knowledge and misconceptions of natural selection. *BioScience*, *57*(3), 263-272.

Nehm, R. H., & Ridgway, J. (2011). What Do Experts and Novices "See" in Evolutionary Problems? *Evolution: Education and outreach*, *4*(4), 666-679.

Park, S., & Chen, Y. C. (2012). Mapping out the integration of the components of pedagogical content knowledge (PCK): Examples from high school biology classrooms. *Journal of Research in Science Teaching*, *49*(7), 922-941.

Park, S., Jang, J.-Y., Chen, Y.-C., & Jung, J. (2011). Is pedagogical content knowledge (PCK) necessary for reformed science teaching?: Evidence from an empirical study. *Research in science Education*, *41*, 245-260.

Park, S., & Oliver, J. S. (2008). Revisiting the conceptualisation of pedagogical content knowledge (PCK): PCK as a conceptual tool to understand teachers as professionals. *Research in science Education*, *38*, 261-284.

Park, S., Suh, J., & Seo, K. (2018). Development and validation of measures of secondary science teachers' PCK for teaching photosynthesis. *Research in science Education*, *48*, 549-573.

Pfennig, D. W., & Ehrenreich, I. M. (2014). Towards a gene regulatory network perspective on phenotypic plasticity, genetic accommodation and genetic assimilation. [Comment]. *Molecular Ecology*, *23*(18), 4438-4440.

Pigliucci, M., Murren, C. J., & Schlichting, C. D. (2006). Phenotypic plasticity and evolution by genetic assimilation. [Review]. *Journal of Experimental Biology*, *209*(Pt 12), 2362-2367.

Sadler, P. M., Sonnert, G., Coyle, H. P., Cook-Smith, N., & Miller, J. L. (2013). The influence of teachers' knowledge on student learning in middle school physical science classrooms. *American educational research journal*, *50*(5), 1020-1049.

Saldaña, J. (2003). *Longitudinal qualitative research: Analyzing change through time*. Rowman Altamira.

Shenton, A. K. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for information*, *22*(2), 63-75.

Shulman, L., S. (1986). Those Who Understand: Knowledge Growth in Teaching. *Educational Researcher*, *15*(2), 4-14.

Sickel, A. J., & Friedrichsen, P. (2018). Using multiple lenses to examine the development of beginning biology teachers' pedagogical content knowledge for teaching natural selection simulations. *Research in science Education*, *48*, 29-70.

Smith, M. K., Jones, F. H., Gilbert, S. L., & Wieman, C. E. (2013). The Classroom Observation Protocol for Undergraduate STEM (COPUS): A new instrument to characterize university STEM classroom practices. *CBE—Life Sciences Education*, *12*(4), 618-627.

Speer, N. M., & Wagner, J. F. (2009). Knowledge needed by a teacher to provide analytic scaffolding during undergraduate mathematics classroom discussions. *Journal for Research in Mathematics Education*, *40*(5), 530-562.

Tekkumru-Kisa, M., Coker, R., & Atabas, S. (2022). Learning to teach for promoting student thinking in science classrooms. *Teaching and teacher education*, *120*, 103869.

Van Driel, J. H., Verloop, N., & De Vos, W. (1998). Developing science teachers' pedagogical content knowledge. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, *35*(6), 673-695.

Wagner, J. F., Speer, N. M., & Rossa, B. (2007). Beyond mathematical content knowledge: A mathematician's knowledge needed for teaching an inquiry-oriented differential equations course. *The Journal of Mathematical Behavior*, *26*(3), 247-266.

Supplementary Materials

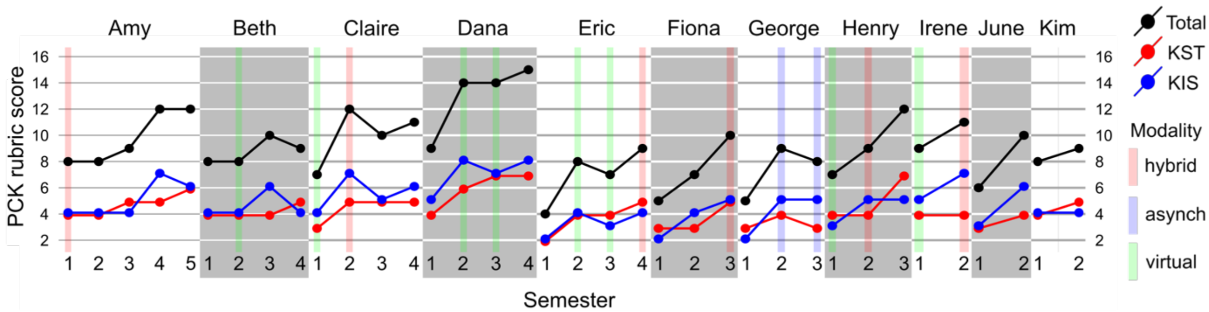


Figure S3.1. PCK rubric scores including information about lesson modality. During data collection, the COVID-19 pandemic forced many participants to change the modality of their focal lesson. We did not detect an effect on PCK scores of lesson modality. Modality of each focal lesson is denoted with colored vertical bars: hybrid (red), asynchronous (blue), and virtual (green). Hybrid lessons were held with some students in the classroom and others joining remotely via teleconferencing software. Virtual lessons involved a full-online, synchronous lesson given via teleconferencing software. Asynchronous lessons involved pre-recorded video lectures. Henry's virtual lesson involved a worksheet which students attempted as homework prior to a flipped-classroom-style virtual session with students. Henry interacted with students' thinking during this synchronous portion. Synchronous, in-person lessons are denoted by the absence of a colored vertical bar.

Table S3.1. PCK Codebook

Using a collaborative and iterative qualitative content analysis, we developed this codebook using four *a priori* codes (bold) for pedagogical knowledge and pedagogical content knowledge. Codes that emerged during constant comparison are denoted in italics.

Code Name	Description
Pedagogical content knowledge of student thinking (KST)	Set of codes that capture evidence of instructors' topic-specific awareness of students' prior knowledge about a topic, common conceptual difficulties, common inaccurate ideas, and how thinking about a topic is likely to change with instruction. In transcripts, this may look like: (1) diagnosing an issue with student thinking, (2) discussing how students typically respond to questions, scenarios, (3) interpretations of typical student thinking, (4) evaluating if their instructional representations made sense to the students, or (5) speaking from the perspective of students. Be wary of cases where the instructors may be talking us through their own thinking about a topic and not the thinking that they believe a student is moving through. This is not KST.
Student validated	Captures PCK-student thinking that is soured directly from their students, comparable students, or research about students' thinking about a specific topic. Students may have provided this information knowingly or unknowingly

	through asking questions of the instructor, responding to exams and activities, and any other student-generated product.
Lacking evidence from students	Captures KST that does not provide any indication that the ideas the instructor has are grounded in actual experience with or evidence from students. This includes instances when an instructor is making assumptions about what students know or will find difficult based on the instructor's own thinking, experiences as a learner, or perceptions of the general public.
Competency specific	Captures knowledge about student thinking that is competency specific. These segments should always be coded with an additional KST code to indicate student-validated or lacking evidence of student validation. This knowledge is about student thinking related to competencies from Vision and Change (AAAS, 2011; Clemmons et al., 2020): (1) apply the process of science (ask questions; analyze & interpret data, plan and carry out investigations; obtaining, evaluating and communicating information; engaging in argument from evidence), (2) use quantitative reasoning, (3) use modeling and simulation, (4) tap into the interdisciplinary nature of science, (5) communicate and collaborate with other disciplines, and (6) understand the relationship between science and society.
Knowledge of instructional strategies (KIS)	Set of codes that captures evidence of the knowledge that instructors are using in making decisions about what strategies to use to teach a specific topic in biology. Strategies may include specific examples they use, analogies, questions, problems, cases, visual representations, explanations, and real-time in class utterances and actions. Often, it is useful to think of evidence of KIS as evidence of an instructor's rationale for any decision they make for the purposes of teaching a specific topic to a specific group of students. We use a liberal definition of what constitutes a "decision." An instructor's discussion of the existence of a practice (analogy, question, etc.) without evidence of their reasoning for this practice should not receive this code.
Tailored	Captures evidence of KIS that is specifically linked to KST. In this type of KIS, instructors give reasoning for their instructional decision that relies on specific insights or knowledge about how and/or why students encounter

	difficulties with a topic. Put another way, the instructor's KST is often inherent to and embedded in their rationale for the design of the instructional strategy. This goes beyond an understanding that students struggle with a particular area of content. Instructors may favorably discuss instructional strategies that force students to confront nuances of the content where students typically go wrong and opt for representations that clarify components of the content that students typically find unclear. In this way, instructors use their KST to inform their instructional decisions, and this means that these two codes (KST & KIS) often occur next to one another or overlap in interviews.
Broad strokes	Captures evidence of KIS that is broadly or vaguely linked to their knowledge of student thinking. In this type of KIS, instructors are drawing on PCK about student thinking, but a shallow level of KST. Most commonly, they know THAT something is difficult for students but not why. This code is for instances where an instructor is clearly making an instruction decision while considering a particular student difficulty in mind, but the decision is not particularly responsive to the specifics of the student difficulty. Therefore, their instructional strategy targets a difficulty but is not actually informed by deep knowledge of the source of the difficulty or the nature of the difficulty. This could be (but is not necessarily) because the instructor lacks those details as a part of their KST. Be on the lookout for decisions that are designed to address a misconception (or lack of knowledge) that are based on wholesale evaluations of class performance rather than the student thinking of individuals that may or may not be evident in the source material of their evaluations. Other times, instructors may have general ideas about why students hold certain misconceptions, but these are often described as perceptions rather than knowledge based on evidence. Plans to address the misconception with broad-strokes strategies reflects this lack of specificity. For instance, an instructor may recognize that a certain concept/task is difficult for students and decide to cover it more in depth, give students more practice with the task, or spend more time with it during lecture.
Lacking evidence of integration	Captures evidence of KIS that lack evidence of a link between the instructional strategy and the instructor's knowledge of student thinking. Instructors are describing the thinking behind a topic-specific instructional decision but the reasoning for making this decision is vague OR

	they have clearly made the decision based on something other than knowledge of student thinking. This does NOT include instructional decisions that are logistical or otherwise based on knowledge instructors have about their students that is not topic specific.
Competency specific	Captures knowledge about student thinking that is competency specific. These segments should always be coded with an additional KIS code to indicate whether the KIS is tailored, broad stroke, or lacking evidence of integration. This knowledge is about rationales for instructional strategies that are used to support student development of competencies from Vision and Change (AAAS, 2011; Clemmons et al., 2020). Competencies are listed in the KST-competency specific code description.
Pedagogical knowledge of monitoring student thinking (MST)[1]	Captures evidence of pedagogical knowledge related to monitoring student thinking during class. This knowledge is generalizable beyond one topic. It deals with the goals/intentions instructors have related to monitoring student thinking, and/or the outcomes that result from monitoring student thinking. These codes may apply to attempts to monitor student thinking during class, before class, or on homework and assessments after class, though in-class is most common in these data.
Pedagogical knowledge of how people learn (HPL)[1]	Captures evidence of pedagogical knowledge related to how people learn. This knowledge is generalizable beyond one topic. It deals with goals, intentions, and rationales related to engaging student in constructing their own knowledge through generative work. These codes may apply ideas about and rationales for integrating generative tasks into lessons and providing students with opportunities to externalize their reasoning. It also includes rationales about the benefits of engaging students in generative work and externalizing reasoning. These codes may apply to any instance where an instructor demonstrates their knowledge of how people learn as they plan, implement, and reflect on a lesson.

[1] These represent categories of codes. We did not use the specific codes within the categories for this research, and instead focused on any knowledge that could meet these criteria.

Table S3.2. PCK Rubric

Directions for using the rubric: The rubric is well-aligned with the codebook. Italicized text indicates code presence and absence aligned with the levels. However, code counts will not be sufficient to use the rubric. Read a full transcript, consider the codes applied, and make judgements about whether the evidence is sufficient to indicate a level of knowledge. Make notes about the specific topics (KST) and instructional strategies (KIS) that you are counting. This rubric captures topic-specific PCK and competency-specific PCK, using the competencies described in Vision and Change (AAAS, 2011; Clemmons et al., 2020). This rubric aims to capture topic- and competency-specific knowledge about student thinking and instructional strategies used in the target lesson. Throughout the rubric the term "topic" is inclusive of topics and competencies. Do not consider utterances about student thinking or instructional strategies outside the target lesson. Do not consider instances of PCK-IS that are hypothetical or plans for the future.

Level of evidence of pedagogical content knowledge *this rubric assumes that knowledge of student thinking and instructional strategies are topic- or competency-specific				
	1: Limited evidence	2: Some evidence	3: Convincing evidence	4: Extensive evidence
Planning (pertains to pre-instruction interview & EXCLUDES anything coded with IS-planning)				
Knowledge evident in anticipation of student's prior knowledge, misconceptions, difficulties with topic(s) taught in target lesson	Some anticipation of student thinking, but no evidence this knowledge is sourced from students *Codes include KST-Lacking evidence from students*	Some anticipation of student thinking, with minimal evidence that this knowledge is sourced from students *Codes include KST-Lacking evidence from students and KST-Student validated about 1-2 topic(s)*	Anticipating student thinking about several topics, and sourced from students *Codes include KST-Student validated about 3-4 topics*	Many instances of anticipating student thinking about different topics, and sourced from students *Codes include KST-Student validated about 5+ topics*
Knowledge evident in planned instructional strategies and their basis in student thinking about specific topics	Some planned instructional strategies, but no evidence they are designed or selected to account for student thinking *Codes include only KIS-Lacking evidence of integration*	Some planned instructional strategies, with minimal accounting of student thinking in design/selection of IS *Codes include KIS-Broad stroke and/or KIS-Tailored about 1-2 distinct strategies*	Several planned instructional strategies that account for student thinking in design/selection of IS *Codes include KIS-Broad stroke and/or KIS-Tailored about 3 distinct strategies (maximum = KIS-Tailored about 2 strategies)*	Several planned instructional strategies that tailor to student thinking in design/selection of IS *Codes include KIS-Tailored about 3+ strategies*
Implementing/Reflecting (pertains to post-instruction interview & EXCLUDES anything coded with IS-planning)				

Knowledge evident in recognition of student's prior knowledge, misconceptions, and difficulties with specific topics during the target lesson	Some discussion of student thinking, but no evidence this knowledge is sourced from students *Codes include only KST-Lacking evidence from students*	Some discussion of student thinking, and minimal evidence that this knowledge is sourced from students *Codes include KST-Lacking evidence from students and KST- Student validated about 1-2 topic*	Several instances of discussing student thinking that is sourced from students *Codes include KST-Student validated about 3-4 topics*	Many instances of discussing student thinking about different topics, and sourced from students *Codes include KST-Student validated about 5+ topics*
Knowledge evident in discussion of how instructional strategies in the target lesson related to student thinking about specific topics	Some discussion of instructional strategies, but no evidence they are designed or selected to account for student thinking *Codes include only KIS-Lacking evidence of integration*	Some discussion of instructional strategies, with minimal accounting of student thinking in design/selection of IS *Codes include KIS-Broad stroke and/or KIS-Tailored about 1-2 distinct strategies*	Several instances of discussing instructional strategies that account for student thinking in design/selection of IS *Codes include KIS-Broad stroke and/or KIS-Tailored about 3 distinct strategies (maximum = KIS-Tailored about 2 strategies)*	Multiple instances of discussing instructional strategies that tailor to student thinking in design/selection of IS *Codes include KIS-Tailored about 3+ strategies*

CHAPTER FOUR

EARLY-CAREER FACULTY PEDAGOGICAL KNOWLEDGE OF HOW PEOPLE LEARN: QUALITATIVE VARIATION, ALIGNMENT WITH LEARNING THEORY, AND LONGITUDINAL DEVELOPMENT[1]

---

[1] Waugh, A.H., K.E. Green, and T.C. Andrews. To be submitted to *CBE—Life Sciences Education*.

Abstract

Active learning has the potential to enhance student learning and other outcomes. However, the effectiveness of active-learning instruction depends on how an instructor implements it, which can vary significantly across different classrooms. Prior studies suggest a link between the effectiveness of active-learning implementation and instructors' pedagogical knowledge, including their understanding of how students learn. Using the ICAP as the guiding framework, this study investigates the variation and development of pedagogical knowledge of how people learn among 12 early-career undergraduate life science instructors who use active learning. We aimed to characterize the alignment between instructors' pedagogical knowledge of how people learn and the empirically tested principles of learning in the ICAP framework. Our second objective was to describe longitudinal trajectories of pedagogical knowledge development to illustrate how pedagogical knowledge can influence the nuances of active-learning implementation. We interviewed participants before and after a focal lesson over the course of several semesters which allowed us to capture knowledge variation and development. Through qualitative content analysis, we identified seven distinct ideas about how people learn and characterized their alignment with the ICAP framework. Case studies of two participants describe how the knowledge development corresponded with changes to their teaching practices. We found that early-career college faculty vary considerably in their pedagogical knowledge of how people learn, and their knowledge can influence important nuances of active-learning instruction. Our results suggest pedagogical knowledge could be an essential link between evidence-based teaching practices and hypothesized benefits to student outcomes.

Introduction

Active-learning instruction has the potential to increase students' ability to learn fundamental concepts and develop scientific thinking skills (e.g., Freeman et al., 2014) and begin to repay educational debt owed to students from historically and currently marginalized groups (Haak et al., 2011; Freeman et al., 2014; Eddy & Hogan 2014; Ballen et al., 2017; Theobald et al., 2020). Though active-learning instruction can be highly effective at improving student outcomes in STEM, it encompasses a wide range of strategies and each can be used effectively and ineffectively, depending on the nuances of implementation (Turpen & Finkelstein 2009, Dancy et al., 2016). As a result, the learning gains that students achieve in active-learning classrooms can vary substantially (Pollock & Finkelstein 2008; Andrews et al., 2011; Nehm et al., 2022). Thus, realizing the benefits of active learning for students depends not just on convincing instructors to adopt these strategies, but also supporting them in using these strategies effectively.

Instructor knowledge is a critical component in effective implementation of evidence-based strategies (e.g., Stains & Vickrey, 2017; Offerdahl et al., 2018) and teaching professional development can help foster needed expertise (e.g., Jackson et al., 2022). Our prior work points to knowledge that may be essential for effective active-learning instruction. Compared to novice active-learning instructors, experts paid more attention to whether instructional strategies provided students with opportunities to construct their own understanding (Auerbach et al., 2018). When we examined the knowledge that active-learning instructors use in their own teaching, those implementing strategies that foster deeper learning relied on the knowledge that students learn by creating explanations of their reasoning (Andrews et al., 2019). In contrast, instructors who primarily engaged students in remembering facts and algorithmic problem-solving focused on using active learning to break up their lectures and practice for exams (Andrews et al., 2019). Together, this work suggests that an instructor's ideas about how people learn relate to the ways they implement active learning in their classroom.

The present study investigated pedagogical knowledge about how people learn among college instructors and how it develops. Very few studies have investigated the development of teaching knowledge among college STEM instructors, and none have examined pedagogical knowledge development (Andrews et al., 2022). Pedagogical knowledge of how people learn encompasses ideas and reasoning about the experiences that foster learning for students and how instructors can create these experiences. This knowledge informs both the design and implementation of lessons and is generalizable across different topics that an instructor teaches (Auerbach & Andrews, 2018). Research in cognitive science offers a strong basis for determining which pedagogical ideas about how people learn are more and less productive for fostering deep student learning.

The ICAP framework, grounded in cognitive science, defines types of active learning and emerged from investigations of the relationships between the cognitive work of students and the learning they achieve (Chi, 2009; Chi & Wylie, 2014). The ICAP framework defines four modes of cognitive engagement: **I**nteractive, **C**onstructive, **A**ctive, and **P**assive (Table 4.1). The ICAP framework has been repeatedly empirically tested across a variety of educational contexts and has proven to be a robust and predictive theoretical framework for understanding the learning experiences that promote student learning (e.g., Chi, 2009; Menekse et al., 2013; Chi & Wylie, 2014; Freeman et al., 2014; Chi et al., 2017; Chi et al., 2018; Menekse & Chi, 2019). The findings of this research, including direct comparisons of learning outcomes for each type of cognitive engagement, support a key principle of teaching and learning: each progressive level of cognitive engagement—passive, active, constructive, and interactive—leads to improved learning outcomes compared to the previous level (I > C > A > P).

For example, students who engage interactively by defending their reasoning in discussions or constructively by creating concept maps develop a deeper conceptual understanding compared to those who engage actively by taking verbatim notes or passively by listening to lecture (e.g., Chi & Wylie, 2014). The work that students do at the interactive and constructive levels of the ICAP framework is collectively referred to as generative work because these levels require students to generate verbal reasoning, written explanations, or other products that go beyond the information provided to them (Chi & Wylie, 2014).

The goal of this study was to characterize variation and development of pedagogical knowledge of how people learn among early-career undergraduate life science instructors using ICAP as the guiding theoretical framework. We focused on early-career instructors because we expected this population to be poised to develop their knowledge and practice as they encountered their first teaching experiences as faculty. Given their lack of experience, these faculty may also be especially likely to seek teaching professional development, making our research findings more directly applicable to educational developers. We aimed to characterize practical teaching knowledge closely intertwined with an instructor's actual active-learning instruction, rather than more general beliefs or philosophies.

This study pursued two research objectives. First, we aimed to characterize the alignment between participants' pedagogical knowledge of how people learn and the ICAP framework, as this theory is robustly supported by evidence about how people learn and highly relevant to active-learning instruction. Our second objective was to richly describe longitudinal trajectories to illuminate possible avenues of knowledge development and nuances of how knowledge can influence teaching. We accomplished the first objective using state-of-the-art qualitative methods to uncover discrete, recurring pedagogical ideas about how people learn that participants used while planning, implementing, and reflecting on their instructional practices. For the second objective, we generated and examined detailed exemplars of knowledge development in two participants, using evidence of knowledge and practices and how they changed over time. This study contributes the first detailed characterization of pedagogical knowledge about how people learn, grounded in theory from cognitive science, and illustrates how this knowledge can develop and influence fine-scale instructional practices.

Table 4.1. Four levels of cognitive engagement in ICAP framework, described by observable student behavior and expected learning outcomes (adapted from Chi & Wylie 2014).

Mode	Student behavior	Learning outcomes
Interactive	Two or more learners discuss and substantively respond to one another to generate outputs that go beyond the information that has been presented in instructional materials (e.g., defending and arguing a position)	Deepest understanding, potential to innovate new ideas, interpretations, products.
Constructive	Learners individually generate outputs that go beyond the information that has been presented in instructional materials (e.g., design an experiment, predict the effect of mutations in a pathway)	Deep understanding, potential for transfer to new contexts
Active	Learners recall information and/or make physical manipulations without adding new knowledge (e.g., taking verbatim notes)	Shallow understanding, potential for transfer to very similar contexts
Passive	Learners receive information (e.g., listening)	Minimal understanding, potential for knowledge recalled verbatim and in identical context

Methods

Participants

We recruited participants who taught large undergraduate life sciences courses (with 50+ students). All participants self-identified as using active-learning strategies and expressed a commitment to further developing these strategies over time. In total, we collected data from 12 participants over a total of 42 semesters of instruction. These participants taught at four public, research-intensive institutions in the Southeastern United States. Each participant held a Ph.D. in a discipline relevant to their teaching. Participants held either a long-term tenure-track (n = 4) or teaching-track (n = 8) faculty position. We recruited early-career faculty for this study, selecting participants who had taught college courses for four or fewer semesters. We collected data from each participant related to a specific lesson in their course, repeating the data collection each semester they taught the lesson over two or more semesters (range = 2-7; mean = 3.5; see Table 4.2). For three participants, we captured data related to two different lessons across the study, but data came from one lesson for most participants (n = 9). Data collection occurred each semester that participants taught the focal course, once or twice annually. Participants taught a range of upper- and lower-division life science courses, including biology, microbiology, developmental biology, evolutionary biology, anatomy and physiology, and neuroscience, catering to both majors and mixed groups of majors and non-majors. The focal courses had an average enrollment of 144 students (SD = 70.0, see Table 4.2). Data collection spanned Fall 2019 to Spring 2023, excluding summers, with the length of participation varying based on when individuals joined the study and when they stopped teaching the focal course (Table 4.2). Due to the COVID-19 pandemic, some semesters of data collection involved remote teaching or HyFlex formats (Figure S1).

Participants in this study stood out from typical life sciences faculty due to their extensive involvement in teaching professional development. During their participation in the study, participants engaged in a range of teaching professional development opportunities, such as teaching conferences, formal teaching fellowships that offered peer-to-peer and professional support, bootcamps to support transition to online teaching, and workshops dedicated to improving classroom diversity, equity, and inclusion. This high level of engagement likely equipped participants with unique knowledge and skills, and suggests high motivation to improve as teachers, which may set them apart from other life sciences faculty. At the study's conclusion, participants completed a survey about their experiences with teaching development, both prior to and during the study. The survey revealed that the majority (9 out of 12) had participated in over 40 hours of teaching development and received formal teaching training during their time as graduate students or postdoctoral researchers. About half (7 out of 12) had engaged in formal teaching mentorships as mentees, and a few (3 out of 12) had even led professional development sessions themselves. Additionally, three participants had published peer-reviewed research on education, and two others had published biology course materials or lessons.

Table 4.2. Participant information.

Pseudonym	# of time points	Course level [a]	Approx. class size	Position	Semesters teaching experience as faculty at first time point
Amy	7	LD	180	Lecturer	4
Henry	5	LD	200	Lecturer	4
Beth	4	UD	80	Assistant Professor	0
Claire	4	LD	100	Lecturer	0
Dana	4	UD	110	Assistant Professor	2
Eric	4	LD	60	Academic professional	2
Fiona	3	LD	75	Academic professional	1
George	3	LD	250	Lecturer	2
Irene	2	UD	200	Assistant Professor	2
June	2	UD	150	Lecturer	2
Kim	2	LD	75	Assistant Professor	0
Lance	2	LD	250	Lecturer	8

[a] UD = upper division; LD = lower division

Eliciting pedagogical knowledge of how people learn

At each time point, we conducted pre- and post-lesson interviews to capture participants' pedagogical knowledge as they planned, taught, and reflected on a target lesson, (Alonzo & Kim, 2016). Interviews also elicited other types of instructor knowledge, which are not the focus of this study. The pre-lesson interview was semi-structured and targeted knowledge used in planning a lesson, such as the design of activities and intended implementation. Prior to the interview, we collected relevant lesson materials such as slides, handouts, and pre-class assignments. Interview questions aimed to explore participants' pedagogical knowledge, asking how specific instructional strategies in the target lesson supported student learning (see full interview protocol in Supplemental Materials Appendices A & B). After the first semester, interviews began with questions about the reasoning behind and changes to the design of teaching materials since the last time the lesson was taught. Pre-lesson interviews occurred one business day prior to the lesson and typically lasted around 60 minutes.

The design of post-lesson interviews incorporated opportunities for participants to view short video-clips of their own teaching from the target lesson to stimulate their ability to recall their thinking during instruction and reflect on the lesson (Alonzo & Kim, 2016). To generate these clips, we video-recorded each target lesson using a camera and a lapel microphone worn by the instructor. These recordings also served as a source of raw data for a systematic analysis of teaching strategies (detailed below). Before each post-lesson interview, we selected three or more clips (range = 3-6 clips) from the recorded lesson. Clips showed students working and, when possible, instructors interacting with students as they worked.

During each post-lesson interview, participants and the interviewer watched selected video clips from the recorded lesson. We asked participants to provide a "running commentary" of what they were thinking during the moments shown in the clips. After they shared their recollections, we asked follow-up questions to delve deeper into their thought processes. Since the content of the clips varied based on the lesson, the interview protocol had to be flexible rather than semi-structured. For each interview, we selected a customized set of questions drawn from a larger pool. We asked each question in the larger pool at least once during each interview, ensuring that interviews were consistent across time and participants despite the variation in questions from clip to clip. The questions included, “In this moment, what were you thinking about your role in this interaction with the student? Why is that role important?” and

"How does this [general instructional strategy, such as think-pair-share, case study, or worksheet activity] support student learning?" (see the full interview protocol in Appendix).

Qualitative analysis of pedagogical knowledge

Our goal was to identify and characterize pedagogical knowledge of how people learn that participants used when planning, teaching, and reflecting on a target lesson, using ICAP as a guiding framework. The ICAP framework outlines student behaviors, not teacher knowledge. To generate *a priori* qualitative codes to capture variation in pedagogical knowledge of how people learn, we drew on our knowledge of the ICAP framework and findings from prior qualitative studies of the pedagogical knowledge of undergraduate instructors (Auerbach et al., 2018; Auerbach & Andrews, 2018; Andrews et al., 2019). We developed three *a priori* codes that distinguished different types of pedagogical knowledge of how people learn at the interactive, constructive, and active levels of the ICAP framework (Chi and Wylie, 2014). Since all our participants used active learning, we opted to allow a code for pedagogical ideas related to how people learn via passive cognitive to emerge if present in our sample. Later, we combined the pedagogical knowledge codes for the constructive and interactive levels into a single code for generative pedagogical ideas, as knowledge related to interactive cognitive engagement was exceedingly rare in our sample. This left us with codes for active and generative work. As we gathered more examples of this pedagogical knowledge, we were able to identify and characterize variation.

Using constant comparison, we developed new subcodes within the original *a priori* codes to capture different features of this knowledge. At least two researchers (A.H.W., K.E.G., and T.C.A.) independently read and coded each transcript, and any disagreements were discussed until a consensus was reached. As we refined the definitions and boundaries of the codes, we reanalyzed all previously coded data to ensure consistency. The result of this process was a finalized codebook (Table S4.2). We developed this codebook during the early stages of the project, using about 50% of the data, and then applied it to the full set of interviews over time for each participant.

Next, we aimed to assess the extent to which participants' pedagogical knowledge of how people learn developed from a focus on the learning benefits of active engagement to a focus on creating opportunities for generative work. This required reducing our rich qualitative data in ways that could be compared across time. To do this, we developed a scoring system based on the specific combination of subcodes present in each interview at an individual time point in the study (Table 4.3). We assigned higher numerical values to ideas about how people learn that aligned with the generative (interactive and constructive) levels of the ICAP framework. We also awarded participants additional points when they demonstrated knowledge of the underlying mechanisms of learning associated with generative work or knowledge of how to facilitate generative work. The quantification system (Table S4.1) was applied to the pre- and post-lesson interview data separately, and the sum of those values represented the participant's total quantified value at an individual time point in the study. We used the presence or absence of a code, not the number of times it was applied to the data, as the basis for scoring, to avoid inflating knowledge values for instructors who share particular ideas repeatedly (Figure 4.1).

Case study selection and analysis

We developed rich cases to characterize the details of how instructors could develop pedagogical knowledge of how people learn and whether and how knowledge development corresponded to changes in teaching. Since pedagogical knowledge development has never been studied in college faculty, we aimed to examine what might be possible rather than to summarize patterns of development across participants. We used a holistic approach to select participants for cases who (1) exhibited a substantial (3+ point) increase in their knowledge value in at least one semester and (2) developed and maintained new pedagogical ideas corresponding to the generative levels of the ICAP framework. These selection criteria left us with three options for case studies: Beth, Dana, and Irene.

For each case, we began by compiling a participant's teaching materials, video-recorded lessons, and all evidence of their knowledge of how people learn from each time point. We wrote summaries of the participant's knowledge at an individual time point, noting which ideas were present, as indicated by which codes had been applied, and also nuances beyond what codes captured. We also documented which teaching materials had changed, and when, as well as whether and how implementation had changed. Coarse-grained observation protocol data (see below) added to our ability to assess teaching practice changes. After compiling evidence of pedagogical knowledge development for each of these participants and writing initial descriptions, we selected Dana and Irene as case study subjects based on illuminating patterns of pedagogical knowledge development, as well as the overall clarity of the participant's quotations for the purposes of reporting. We continued to refine our case descriptions of Dana and Irene's knowledge and practice development until two researchers (A.H.W & T.C.A) agreed that the case description was complete, accurate, and accessible to a wide audience.

Trustworthiness of these qualitative analytic approaches

Trustworthiness in qualitative research is established through four key domains: confirmability, dependability, credibility, and transferability (Guba, 1981; Anfara et al., 2002; Shenton, 2004). In this study, we incorporated specific practices throughout data collection, analysis, and reporting to ensure a trustworthy approach to our qualitative work.

Confirmability and dependability are critical in maintaining the integrity of qualitative data interpretation. Confirmability refers to ensuring that interpretations accurately reflect the original data, while dependability focuses on the consistency and repeatability of the findings by different researchers (Anfara et al., 2002; Shenton, 2004). To achieve this, we employed constant comparison methods, with multiple researchers deeply engaging with the data. By requiring all interpretations to be discussed collaboratively and reach consensus, we mitigated the risk of individual biases influencing the outcomes. This process ensured that our analysis was not driven by any single researcher, enhancing both confirmability and dependability.

Credibility, another domain of trustworthiness, refers to the alignment between the research focus and the methodology, ensuring that the study measures what it intends to (Shenton, 2004). Our data collection methods were grounded in established literature on pedagogical knowledge, using well-documented approaches such as stimulated recall (Alonzo and Kim, 2016). These methods were specifically designed to explore early-career instructors' pedagogical knowledge in planning, enacting, and reflecting on instruction, ensuring that the data accurately addressed our research questions (Andrews et al., 2019).

Lastly, transferability involves assessing the extent to which the findings can be applied to other contexts (Shenton, 2004). While we do not claim generalizability due to our sample size and participant diversity, we enhance transferability by providing detailed descriptions of our participants, their teaching environments, and their institution types. This level of transparency allows others to assess the relevance of our findings to similar contexts. Additionally, we use precise language to limit our conclusions to the study population, acknowledging the specific context and scope of our research.

Analyzing teaching strategies

We conducted a systematic analysis of participants' teaching practices during each target lesson, focusing on four specific behaviors: instructors interacting with students in small groups or one-on-one, students working individually or in small groups during class on active-learning assignments, and students explaining their reasoning to the instructor. The first three behaviors were adapted from the Classroom Observation Protocol for Undergraduate STEM (Smith et al., 2013), while the fourth was developed to capture instances when the instructor heard student reasoning. We defined students explaining their reasoning as any instance where a student verbally elaborated on their thinking to the instructor. This could happen in various formats— one-on-one, in small groups, or in front of the entire class—as long as the instructor appeared to be listening. To ensure we captured elaborations, we only counted explanations that involved more than a single word or phrase.

To quantify these behaviors, we coded each two-minute segment of the video-recorded classes, marking the presence or absence of each behavior. We then calculated the average percentage of segments containing each behavior across the target lessons for the instructors featured in case studies. A coding team, trained as part of a larger project, followed a rigorous protocol to ensure consistent analysis. To assess inter-rater reliability (IRR), we used Cohen's kappa. Once high IRR was achieved, individual coders analyzed the remaining videos, recalculating IRR periodically to maintain consistency over time. Of the 42 videos analyzed for this study, 17 were reviewed by multiple coders, resulting in an average Cohen's kappa of 0.96, indicating strong agreement among coders.

Results

We present results in three sections. Accomplishing our first research objective, the first section characterizes qualitative variation in participants' pedagogical knowledge by describing seven recurring, discrete ideas about how people learn and how they align with the active and generative levels of the ICAP framework. The second and third sections of the results address our second research objective which was to describe longitudinal trajectories of pedagogical knowledge development and nuances of how knowledge can influence teaching. The second section presents our longitudinal, quantitative assessment of pedagogical knowledge development. The quantitative data reveal patterns over time and across participants, but it reduces much of the important complexity in how participants' knowledge developed. The third section presents two in-depth case studies of pedagogical knowledge and teaching practice development over time. Unless otherwise indicated, all text within quotations are verbatim quotes from participants, with minor editing for clarity.

Qualitative variation in pedagogical knowledge of how people learn and alignment with ICAP

Participants relied on a range of ideas about how people learn as they planned, enacted, and reflected on active-learning lessons. Each section below describes a pedagogical idea about how people learn that was common among participants, starting with ideas more aligned with the passive and active ends of the ICAP spectrum and proceeding to ideas about generative cognitive engagement. Our qualitative analysis captured participants' knowledge about their teaching practices, not the practices themselves. Most of these pedagogical ideas dealt with the teaching strategies that participants choose to employ (Table 4.3). Two ideas, however, stood out from the rest. One pedagogical idea, which we termed "generative work as a mirror," encompassed an idea about the mechanism by which generative work fosters student learning. Another pedagogical idea, which we refer to as "refraining from explaining" emerged when instructors described their intentions as they facilitated student work in real-time during class.

Activity not passivity maintains students' attention and moves beyond memorization

Since all participants used active-learning strategies in their course, it was not surprising that they had ideas of the advantages of active learning over traditional lecture. Participants explained that they used active-learning strategies because they created opportunities to break up lectures, which they thought would help sustain students' attention. For instance, Eric asked a discussion question near the beginning of class and explained the advantage of this strategy this way, "I think I've talked for about 5 or 10 minutes here. This [question] is just to get them engaged, right? To break up the time of me talking and droning on." Participants commonly expressed this idea. In some cases, they seemed to hold an underlying assumption that students learn by listening to lectures but needed breaks so that they could continue to pay close attention during lecture. Thus, their knowledge about how people learn framed active learning as a tool to keep students more engaged during lecture, positioning passive cognitive engagement as the most important learning opportunity in class, a view out-of-line with the ICAP framework (Chi & Wylie, 2014). Participants also used active-learning strategies because they wanted students to do more than memorize information. Kim explained, “maybe this is a little controversial, but I don't see any value in memorizing material. I memorized the Calvin cycle on like four different occasions over the course of my education and promptly forgot it immediately afterwards, until I needed it. And we live in an age where you can look up anything in an instant. So, I no longer really think that the important thing we're doing here is getting them to understand the Calvin cycle. Instead, I think, in an age where you can Google any fact in a minute, what's much more important is to be curious so that you do look up that information and then to know what to do with the information once you have it.” Similarly, Amy aimed for students to do more than passively receive information. Rather than give students a table of completed information, Amy asked students to fill in the table with their predictions, “if I had shown them this table without giving them a chance to fill it in themselves, there would've been no reason for them to think about why this happens or how this actually looks in reality. They just would've copied it down and then memorized it.” Here, Amy conveyed a common idea among participants: passive delivery of information to students would not achieve the type of learning she hoped to foster in their course.

Peer-to-peer listening allows students to hear ideas from each other

Every participant engaged students in some form of small-group work during class, and some participants valued group work primarily or solely because it created an environment in which students could learn from their peers (Table 4.3). For instance, Amy explained that students were “all coming with different strengths as far as biology and that stuff goes. So, by giving them a chance to talk to each other, the ones that have more advanced prior knowledge are able to help those other folks along.” Here, Amy is describing how group work fosters learning because students can learn from more knowledgeable peers. However, it is unclear if the cognitive engagement of the students with less advanced prior knowledge goes beyond passively listening. Some participants thought students’ peers could offer different and potentially more relatable explanations than the instructor. When asked to explain how talking to each other helps students learn, George responded, “I think they hear a different voice. I think it's that simple. I think hearing it from somebody else who's learning it [for] the first time, they might have a different way of explaining it or thinking about themselves that might resonate with one of them.” Overall, this pedagogical idea prioritizes students listening to each other, which is passive engagement, but some participants also had additional ideas about how group work could foster learning.

Struggle makes it stick because difficulty focuses student attention

At times, participants discussed the relationship between the difficulty of a task and its effectiveness for student learning. Often, instances of this idea lacked an associated underlying mechanism to explain why difficult tasks support student learning. However, when participants described an underlying mechanism, they usually explained how struggling focuses students’ attention and facilitates later recall. When participants displayed these pedagogical ideas, they explained that they designed active-learning opportunities to expose students to challenging tasks. For instance, Lance (Lance) incorporated a difficult question in which students calculated the membrane potential of a cell, and described it this way, “It's hard to put together. I understand that. I want them to struggle through this so that then, you know, we talk a little bit about it in class and when they’re doing their own learning after class, they can say, ‘Oh, I remember thinking about that specifically’ and not just, ‘Dr. [Lance] talked about it’… It's the way we learn everything. Learning is always in the struggle… Of course, every research study ever done is that hearing someone talk about something is not nothing, but it's not as helpful as us digging into it ourselves.” Here, Lance expresses the idea that engaging students in work that involves some struggle will foster greater learning than if students listen to him explain the content. One way in which he expects this to foster learning is that students will pay more attention to a particular topic later when they study because it was difficult for them in class. Interestingly, this perspective positions the learning benefits of active learning as occurring outside of class, as well as in class. As another example, Beth explained, “maybe having some wrong answers in class will inspire some of them to get that information pulled out and together before the test.” In general, the idea that struggle fosters learning hints at, but falls short of the idea that generative cognitive engagement fosters greater learning because the nature of the work asked of students was typically unclear and the main pedagogical goal was to focus students’ attention on particular content.

Application and practice foster deeper learning and skill development

Participants often used active learning because they wanted students to apply their knowledge to novel scenarios and practice using skills. Since they expected this level of mastery on summative assessments, they used class time to engage students in this work. Sometimes participants aimed for students to practice using terminology and definitions. Fiona’s lesson involved an opportunity for students to practice distinguishing between two related concepts in a real-world scenario, “I think it's helpful because, you know, sometimes we just need practice with those vocabulary things… [Students] may feel like they know it in the moment and then, when we're asking them to apply it for the exam, they're not able to because they haven't practiced that way. They haven't practiced that application-type question.” Other times, participants explained that particular concepts or skills required repeated exposure and practice to develop fluency, and they used active learning to offer that practice to students. For

Dana, in-class activities offered opportunities for students to gain additional experience, in a low-stakes environment, applying their knowledge in a complex, novel scenario, “part of the logic of these assignments is modeling the thinking that really they should be able to do for assignments and tests and things like that. So, they have a couple of these questions on exams and part of this interaction is you have practice with it. It is not something that is super intuitive. It does take a little bit to get your head around it. So, now it's your low consequence — talk through it with me, or you have a couple of times to go through and see if you got it right, before it kind of comes to a more stressful time to where there's an exam grade riding on it.”

Thinking beyond exam preparation, some participants viewed opportunities for application and practice as essential for learning. For instance, Fiona explained how chances to apply their knowledge both prepared students for exams and fostered the type of thinking abilities that students need as people and professionals, “We've very much moved away from just regurgitation of knowledge on the exams and more into the application. As a practicing professional, you will have the ability to look things up. That is the reality. You won't have the thinking skills unless we really help you develop those. And so that application is for both. Can you apply what you just learned and hopefully remember and take it with you? But, also practice applying what you learned and thinking through and reasoning out the different scenarios because that's really what the long term goals for the program are. That's what a lot of people hope to gain after a university education, and so it aligned better.” Though participants described different goals for engaging students in applying knowledge, from accurate use of terminology to problem-solving, these ideas often aligned with engaging students in generative work, as defined by the ICAP framework (Chi & Wylie, 2014).

To explain is to learn: outward expressions of reasoning foster deeper learning

Participants offered the most robust evidence that they prioritized generative cognitive work when they articulated that students learn the most from expressing their reasoning about a topic, typically via verbal explanations or written responses to questions. Frequently, this pedagogical idea emerged when participants shared their rationale for using group work. Kim talked about how the process of explaining one's reasoning during group work is crucial to learning, "[Students] can take on a teaching role and that can help their own learning. I think the best way you learn something is by explaining it to somebody else. So, I'm thinking that can help them feel challenged. Like they're being asked to really explain it to somebody else." This common idea was clearly aligned with generative work, in contrast with the other main rationale for group work, that students could benefit from hearing ideas from peers.

Participants also described the learning value of students expressing their reasoning in individual work. For instance, Beth implemented a short writing assignment in the middle of class, and described how this helped students learn, "I think when they have to write down an answer to the question, I think that helps them to be focused on a task to completion. You know, if I ask them to think about something, they can think about it, but them thinking about it doesn't really look any different from them planning their grocery list or whatever. When they have to write, they're pretty engaged in the task and they're gonna finish their answer and writing is a little slower than thinking. And so they're going to have to go through a thought process, whether that thought process is correct or not." Here, Beth distinguished between the work of outwardly expressing their thinking versus a strictly internal mental exercise, emphasizing that explaining their reasoning forces students to engage more deeply with the material. Typically participants who expressed this pedagogical idea wanted to create these opportunities in order to foster deep conceptual learning, but sometimes they described how explaining one's reasoning helped students recall content. As was true for the pedagogical ideas described above, participants often did not share how explaining one's thinking fosters learning, but, as we outline in the next section, some did describe a mechanism.

Generative work as a mirror creates opportunities for students to realize what they do and do not know

Among participants who displayed the knowledge that generative work fosters deeper learning than passive or active engagement, some articulated a mechanism for this learning. These participants explained how generative work allows students to realize what they do and do not know. For instance, June contrasted how students tend to evaluate their own understanding favorably when they hear an explanation from someone else, but then realize the shortcomings in their thinking when they are forced to explain to someone else, “I think it's easy to feel like you understand something when you hear it. Often, you don't realize the things you don't understand until you try to explain it to somebody. I feel like that all the time where I'll start teaching something and I'll realize like, ‘Oh, there's some nuance to this.’” June reflected on her own experience recognizing the limitations of her knowledge as an instructor and aimed to leverage those insights to create similar opportunities for her students.

As another example, Amy explained the learning value of asking students to make a prediction by drawing a graph of anticipated results, “I think it's just like concrete…once their pencil hits the paper, even if they think they understand it, just kind of listening to it, they can more easily move on without actually coming to terms with that in their own head. But if I'm asking [them] to actually draw something on their paper, it's in that moment that they're like either they know it or they don't, they're gonna draw something or they're not gonna be able to. And that's either a wake up call that they don't necessarily understand what it is I'm asking or, or what this concept is, or they do and they can make that connection.” Most commonly participants linked the pedagogical idea that generative work acts as a mirror to opportunities they created for students to explain their reasoning. However, a few participants linked this mechanism to opportunities for application and practice, explaining that practicing a skill or attempting to apply a concept reveals to students what they do and do not know.

Refraining from explaining keeps the onus on students engage in generative work

In a few instances, participants described their responsibility, as facilitators during class, to maintain the opportunities they had created for students to engage in generative work. They recognized that their own facilitation could undermine the cognitive demand required of students in their carefully designed, generative tasks. Participants described how they had to refrain from offering their own explanations and instead prompt students to explain their reasoning. For example, Beth described how she consciously avoids providing answers to student questions, “I almost never give students the answer. I will do something like kind of nudge, nudge, nudge, and then when they get close to the right answer, I won't even necessarily stick around to acknowledge it. I'll say like, I feel like you've really got a handle on this. I'm gonna walk away. So you can finish making this connection yourself and then call me over if you need more help. So I like the idea of them getting there on their own steam, and I know they can do it.” Beth explained how she will often leave an interaction with students when they are close to arriving at the correct understanding because she wants students to make connections on their own.

Similarly, participants explained how they entered whole-class discussions with the intention of letting student ideas chart the path forward. Eric described how he restrained himself as he facilitated a discussion in his class by thinking, “let the students drive it, right? I know what the answer is, and I'll show the answers that I was thinking of in a bit, but I don't want to constrict them too much from the start because when they come to the conclusion on their own, then it's more cemented… So trying to be as hands off and more of a guide in the right direction.” Eric thinks that students will learn more by arriving at an answer on their own, and therefore eliciting student explanations must be prioritized over providing his own explanation.

In other cases, participants described maintaining the cognitive demand of their in-class tasks meant by withholding particular information, leaving room for students to independently connect concepts or apply knowledge from another lesson. Henry did this when he intentionally omitted information from his introduction to a task that students would need to solve a problem.

Henry did so to ensure that, when it came time for his in-class task, students had the opportunity to apply to a novel scenario a concept they had already encountered in other lessons, "I never expressly said what direction the anticodon would have to be in during my explanation. And that was just to see if they could figure out on their own that, like in all previous cases, the two have to be in opposite directions. So I was hoping that they could put that together." Henry described that creating conditions for generative work requires thoughtfully designed tasks that pinpoint important conceptual leaps and deliberate facilitation that specifically avoids undermining students' opportunities to learn by making those conceptual leaps for themselves.

Table 4.3. Pedagogical knowledge about how people learn participants relied on as they planned, enacted, and reflected on their active-learning instruction, with example quotes.

Pedagogical ideas about how people learn	Example quote
Active rather than passive: opportunities to do active work in the classroom yields better learning than passively listening to lecture (*Active*)	"If you're just listening, you're passively learning. If you're writing or talking or gesturing, I think those are signs of probably more engaged learning… I think those activities help them become more active in their learning." – Beth
	"The fastest way to prepare for me is just traditional, boring lecture using previous PowerPoints that other people have designed. But I find that the students are not asking questions and they're really confused, [and] aren't able to apply that content to the next lecture. And so when I started adding in these break points to the lecture to do little activities or get feedback from them on and talking to their neighbor and reporting back, I found that they were more engaged." – June
Peer-to-peer listening: learning occurs when students hear ideas from peers (*Active*)	"I suppose it's almost similar to a think-pair-share exercise in terms of that they then get to talk it through. They get to hear other people's way of thinking about this. It gives them an opportunity to update their thinking based on what their neighbors are thinking." – Irene
	"So, I think my hope with the groups was that it would provide some knowledge sharing between the students" – Kim
Struggle makes it stick: learning occurs	"I feel like if you struggle with something and get it wrong, then you might be more likely to kind of work through that instead of I just

when students remember what they struggled with during a difficult task (*Active*)	tell them this and then later they're like, 'How do I remember? I don't know.'" – Beth "I think my goal in designing it is— I do think that when you're asked to think about something hard first, before you get any information, it really that then it sticks. Like you've created a framework or context that then you can slot that new information in." – Kim
Application and practice: learning occurs when students practice skills and apply their knowledge to novel scenarios (*Generative*)	"You get a little immediate practice because, too often, I think we explain something in class, class ends and then we say, go do your homework. No one does their homework the first day. You know, I never did my homework like immediately, and by the time you come back to it, you've forgotten most of it. Using the clicker gives them a little bit more review to kind of reemphasize, 'Here's the term we went over, it's important. I want you to get out the pencil and paper and do the math, you know, kind of build that muscle memory'". – George "This class is very application focused. It's all about applying information, not just knowing it. So, um, if you approach the class from, you know, reading the textbook and just watching videos, or just going to class. That, because I know this information that I understand it, that's where the struggles come because you don't really understand it. You're very familiar with it, but you don't really understand it because you can't apply it." – Henry
To explain is to learn: learning occurs when students outwardly express their thinking, often via verbal and/or written explanations (*Generative*)	I always think about [student-led learning] in terms of anything that the students are doing, where their voice is the one sharing the knowledge. In this case, they're sharing… It's giving them the opportunity to explain. I feel like every time that I truly know something is when I can explain it well and someone else understands. Giving them that opportunity to do that might help them realize that they also know it by explaining it. – Fiona "I prefer that they work on them together with other students because it generates more outward expression of what you are thinking. So the students, rather than just being an internal dialogue with yourself, as far as what you're doing, when they're working in groups, they have to explain to each other what they're thinking, bounce ideas off of each other." – Henry
Generative work as a mirror: learning occurs when students reflect on their own thinking as a result of engaging in generative work (*Generative*)	"... sometimes our students don't know what they don't know and they're really clever. They're really bright. If they feel like they recognize things, they tend to think they know more than they did. But when we give them this opportunity that's just a little bit more challenging and they continue to be challenged by it, that might be that realization, 'I actually don't know this as much as I thought I did.'" – Fiona

	"Instead of just listening and digesting, now they're applying and making predictions… And then, if they predicted wrong, then maybe they can think about, 'Okay, what was my line of thinking here and why was that wrong?' So it's getting them to be a little bit more actively participating with what's being presented to them" – Claire
Refraining from explaining: learning occurs when instructors keep the onus for cognitive work on students by prompting students to articulate their reasoning rather explaining answers (*Generative*)	"I have to make a kind of quick decision about, 'Do I just answer the question or do I ask more of the student?' Sometimes when students ask a question that's maybe a little bit more complicated, I'll draw out more or ask the class, 'Think about this for a little bit. That's a great question. What have we learned about that?" When it's a question like bone tissue versus marrow, that one the student should really know and there's not much more to explain about it, … I'll just give them some assurance… So sometimes [my role is the] assurance giver: 'Yes, you're thinking about it right', but sometimes it's, 'I'm not going to answer that question for you, let's think through that.'" – Lance
	"I'm trying to identify how far through this thought process or through this skill they're able to navigate on their own and then see how little I can give them to go the rest of the way…If they just need a tiny little push or a tiny little point in a direction, that's making them use a lot more brain power and a lot more of these skills we're trying to build than if I kind of told them they were wrong and then told them the right answer. So, giving them as little as possible to just kind of nudge them down the road, get them started again is what I hopefully try to do all the time." – Amy

Development of instructors' pedagogical knowledge of how people learn

Our longitudinal design allowed us to track how participants' pedagogical knowledge of how people learn developed over multiple semesters and how this development corresponded with changes to their teaching practices. First, we present longitudinal trajectories of pedagogical knowledge development generated by applying our quantification system to interview data at each time point in the study (Figure 4.1; Table S4.1; see Methods). These trajectories show how participants' knowledge changed dynamically over the longitude of the study.

Longitudinal quantification of pedagogical knowledge development

By quantifying participants' pedagogical knowledge at each individual time point, we reduced the complexity of our large qualitative data set which enabled us to systematically assess changes in participants' knowledge across time. We quantified participants' knowledge of how people learn using based on the specific combination of subcodes present in each interview at an individual time point in the study. Higher values correspond to knowledge more consistent with the higher levels of the ICAP framework. Therefore, an increasing value over time would suggest a participant is developing pedagogical knowledge of how people learn that aligns with ICAP's generative levels of cognitive engagement.

However, participants did not show clear patterns of developing pedagogical knowledge that aligned with the generative levels of the ICAP framework over time (Figure 4.1A). Some participants exhibited a large single-semester increase, a few exhibited a more gradual developmental trajectory, and still others showed an overall decrease in their value by the end of the study. Seven out of twelve (58%) participants increased their value by more than one point by the end of their participation in the study. This increase in value corresponds to the addition of one or more of these pedagogical ideas: refraining from explaining, application and practice, the role of generative work as a mirror, and to explain is to learn. Some participants exhibited more stability in their knowledge than others, meaning that once they developed a new idea, they continued to display it in subsequent interviews (Figure 4.1B). Other participants exhibited less stable knowledge development. These data also reveal that participants vary in terms of their overall level. Some participants, like D8, started the study exhibiting knowledge aligned with generative levels of the ICAP framework, whereas other participants did not. Overall, these data suggest that pedagogical knowledge development does not necessarily occur in the natural course of teaching, even when instructors participate actively in teaching professional development.

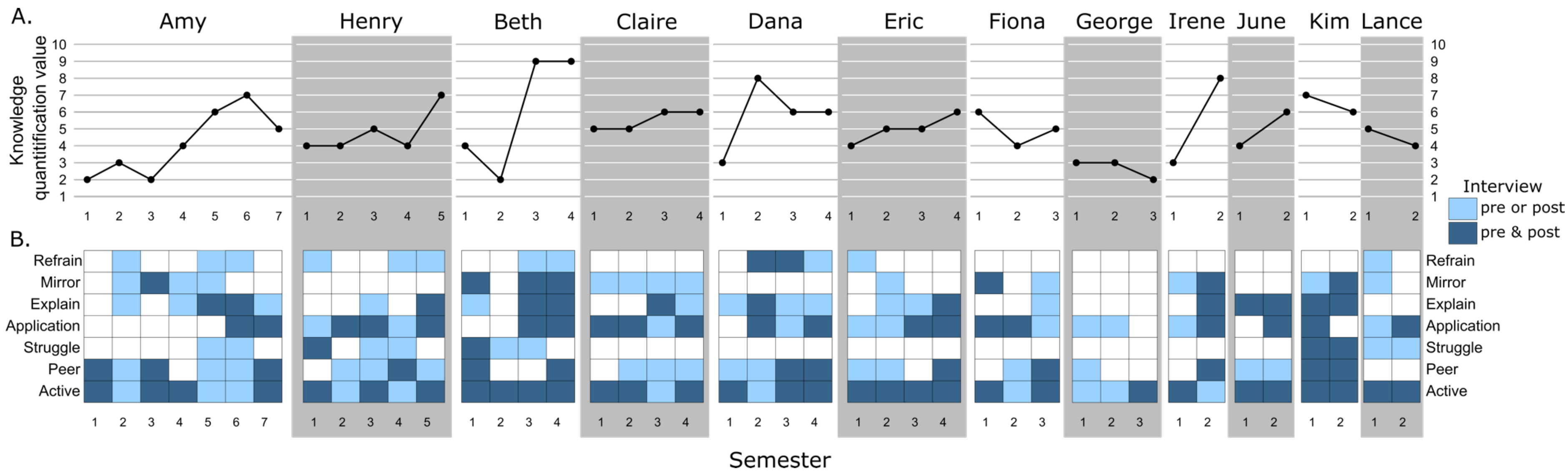


Figure 4.1. Longitudinal trajectories of pedagogical knowledge of how people learn, by participant. (A) Knowledge quantification yielded values to measure participants' pedagogical knowledge at each time point by applying the same system to the specific combination of codes present and absent in the pre- and post-lesson interviews separately. The quantified value represents the sum of values from the pre- and post-lesson interview is shown. Data include each semester an instructor taught the target lesson over a two- to seven-semester period. Min and max values are two and ten (see Methods). (B) A tile plot showing the presence and absence of each code from Table 4.3 at each time point. A word from each code is used as abbreviation: Refrain = Refraining from explaining, Mirror = Generative work as a mirror, Explain = To explain is to learn, Application = Application and Practice, Struggle = Struggle makes it stick, Peer = Peer-to-peer listening, Active = Active rather than passive. A value (A) at an individual time point can be calculated by identifying the specific combination of codes present in each interview (B) using the scoring system given in Table S4.1.

Case studies of pedagogical knowledge development and instructional practice change

We present two exemplar cases of knowledge development, with the goal of uncovering ways in which this knowledge can develop and its potential interplay with teaching practices. We selected participants for case study based on two criteria. Candidate case study participants (1) exhibited a substantial (3+ point) increase in their knowledge value in at least one semester and (2) developed and maintained new pedagogical ideas corresponding to the generative levels of the ICAP framework. The first case study describes how Dana developed knowledge about refraining from explaining and began consistently eliciting student reasoning during the focal lesson. The second describes how Irene developed knowledge that to explain is to learn before she linked this idea to generative work as a mirror for students. Irene then implemented teaching practices that created more robust and frequent opportunities for this kind of cognitive work in the focal lesson.

Case 1: Dana develops and implements pedagogical knowledge about refraining from explaining

In her upper-division developmental biology course, Dana used a flipped classroom approach in her focal lesson on the mammalian sex determination pathway. Dana called these flipped classroom lessons "pathway days," and she used them as a counterpart to her non-flipped active-learning lessons to "really delve into" something from prior lessons so students "can work through it step-by-step and understand something that is yes, a little bit complicated or maybe something that is an offshoot, but enforces other concepts that we've learned through the semester." In the target lesson, Dana gave students a schematic of the steroid hormone biosynthesis pathway showing how, step-by-step, various enzymes facilitate the production of hormones and other molecules important for sex determination. During class, students worked through eight multiple choice questions. Neither the questions nor the pathway schematic changed at any point during our four semesters data collection. These multiple-choice questions varied from engaging students in generative work (e.g., identifying probable outcomes of perturbations in the pathway) to recalling processes.

In all four semesters, Dana interacted with students for more than 90% of the target lesson as they worked in small, informal groups. Despite interacting one-on-one with small groups for most of the class, students explained their reasoning to Dana in less than 10% of two-minute lesson segments until the final semester. In the first three semesters, Dana primarily asked students to recall information when she talked to small groups, using questions like, “What hormones do the testes make?” and then explained her own logic for how to work through the pathway to answer questions (see examples of these interactions from each semester in Supplementary Table S4.3).

In her first three semesters, Dana’s approach to small group interactions was at odds with her pedagogical knowledge of how people learn. Dana expressed that opportunities for students to explain their reasoning created important moments for learning, “Forcing them to talk it out, I think, is useful because, even talking to yourself, going with the idea of like, ‘You don't know something until you have to teach it’... I think just forcing them to put their logic in words. I think you can kind of tell your story in your head, but when you say things out loud—like when we say, ‘Oh my gosh, I just heard what I said.’ It's that type of processing that you're not only saying it but you hear it at the same time. I think just talking it through out loud can be very helpful. Even if you're just talking to yourself.” However, as Dana interacted with small groups, her actions ushered students to the right answer rather than eliciting their reasoning, thereby missing opportunities to let students learn by explaining their reasoning to her.

In the second, third, and fourth semesters, Dana displayed a new pedagogical idea: facilitating generative work necessitates that she refrain from explaining in order to keep the onus on students to explain their reasoning. Dana described her strategy when working with small groups this way: “as much as possible, I try to ask questions.” She also reflected on the tension she experienced as she taught between explaining to students and creating space for students to explain their thinking to her, “I'll admit, that's something I'm trying to be better about… I think sometimes it's definitely easier to just, ‘Let me just explain this to you, instead of really making you work through it.’” Dana pointed out that her role as an information resource in the classroom is a double-edged sword during pathway days because “for some students, I think it would be a beneficial thing to be able to grab me and… clear something up a little more quickly. For other students, I think I might serve more as a crutch.” Dana was aware that students sometimes needed explanations but also recognized that students might use the instructor to bypass cognitive work important for their learning during class.

Dana discussed how resisting the temptation to explain is challenging when students’ reasoning is incorrect but that showing this kind of restraint was crucial for her student’s learning. Commenting on a longer than usual pause after she asked an engaging question to a group of students, Dana explained, “one of the challenges with this is trying not to talk, trying to wait it out. Because again, I can say it all I want, but I'm not the one who needs that practice verbalizing it. I already know what I'm talking about.” Dana implied that sometimes remaining silent is the best instructional move a teacher can make. An important aspect of Dana’s thinking that seemed to make her comfortable with waiting for students to respond was her understanding that, as a content expert, she did not need to explain her reasoning. Instead, she thought these opportunities for learning should be left for students.

Alongside this new idea, Dana described changes she planned to make to her teaching practices. She aimed “to prompt [students] to explain your thought process on this. And I think I do a lot more of, you know, ‘Well, why is that? Like, why do you think that, not that it's right or wrong. You're right. But just explain that to me so I can try and get at what the process is, not just the endpoint.’” Dana also continued to experience tensions between her instructional goals and perceived limitations, “I will admit it. I don't always do that. Sometimes, I launch into explanations because other people are waiting or things like that. It's just the time logistics. But, in an ideal world where I have the time, it's ‘Okay, talk me through your answers. Why do you think this one's correct? Why do you not think that one is correct?'" During class, Dana used her planned lead-in prompt, "Talk me through your logic" a handful of times, which successfully elicited students' reasoning. However, after this lead-in prompt, Dana usually proceeded to do most of the explaining for the remainder of the interaction, asking students only basic recall questions (see Figure S1). Systematic video analysis of her instructional practices indicated no change in how often she heard students' explanations in the second or third semester.

In the fourth and final semester, Dana continued to demonstrate knowledge about refraining from explaining, and her teaching practices changed significantly. Dana heard students explaining their reasoning in roughly 50% of two-minute lesson segments, which was the highest percentage of class time spent hearing reasoning across all participants and all time points in the study and a five-fold increase from Dana's prior target lessons. Looking more closely at her interactions with small groups, Dana often refrained from explaining and instead facilitated generative work by asking questions. Dana frequently took time to ask students to clarify what they had said or to restate their thinking in a different way. Dana prompted students to make predictions and when students provided their prediction without accompanying reasoning, Dana followed up to ask, "Why?" These nuanced behaviors during small-group interactions marked a departure from semesters one, two, and three, where Dana did the bulk of the cognitive work by asking pointed recall questions to walk students through the logic needed to reach an answer. In semester four, Dana kept the onus on students to do the generative cognitive work.

Even though Dana developed ideas about facilitating generative work in her second semester, we did not see Dana implementing this knowledge as part of her teaching practices until semester four. We do not have causal evidence to explain why this development in her practices occurred later, but the sequence is consistent with a scenario where this knowledge development informed changes to her practice. We may have observed a delay in timing between the pedagogical idea and the practice for a number of reasons, including the fact that the lesson was held virtually in semesters two and three. An alternative explanation could be that it simply took time to develop the skilled practice of consistently eliciting student reasoning during class.

Case 2: Irene develops and implements her understanding that to explain is to learn as she refines her teaching practices and materials

As part of her upper division evolutionary biology course, Irene taught her target lesson on sources of phenotypic variation (i.e., genes, environments, gene by environment interactions). Safety precautions during the COVID-19 pandemic necessitated virtual, synchronous instruction in semester one and a hybrid format in semester two, with half of students attending in-person and the other half attending remotely. During the target lesson, Irene alternated between lecturing and giving students opportunities to work on active-learning assignments, including open-ended and closed-ended questions (e.g., multiple-choice, true-false).

Irene's students worked in roughly 25% of two-minute lesson segments in both semesters. In the first semester's virtual lesson, Irene did not interact with students one-on-one, mostly likely due to the challenges of virtual instruction, and in the second semester's hybrid lesson, she did so in 6% of the two-minute lesson segments. In both semesters, Irene did not hear students explaining their reasoning at any point in the lesson. Though these coarse-grained measures of instructional practices did not change across Irene's two lessons, Irene made important changes in the design and implementation of her active-learning assignments. As we describe below, these changes aligned with her developing pedagogical knowledge of how people learn via explaining their reasoning.

In semester one, Irene relied on ideas about active learning benefiting students through opportunities to practice applying knowledge and activity being better than passivity and did not express the pedagogical idea that students learn by explaining their reasoning. Irene shared that “being an active participant in the process and applying is just a much better way of learning. They're applying their knowledge and sometimes the questions are simple in terms of — they just have to regurgitate… But most of the time I'm trying to get them to apply.” Irene explained that a teaching professional development workshop from her time as a postdoctoral researcher had influenced her decision-making on which types of activities to include as active-learning assignments. Irene explained, “I suppose the one thing that I always took from the [organization name] workshop was that you can, even if you've got a big class and multiple-choice questions, you can ask those higher learning things you don't just have to [ask basic recall questions].” Irene also expressed the idea that generative work could provide a mirror for students as she described students interpreting graphs and “think[ing] critically about the topic,” “I suppose, from things that I've read on my own and personal experience, doing is better than listening and that is where things start to make sense. Or, you realize they don't make sense and then you can ask for help. It’s to get them to realize if they are understanding or if they're not understanding.”

In the second semester, Irene expressed a new piece of pedagogical knowledge: students learn when they explain their reasoning. This development in her knowledge coincided with changes in how she implemented active-learning assignments in the focal lesson, as well as the questions she posed to students on slides and verbally. During the target lesson in the second semester, she encouraged students to talk in small groups before answering poll questions. Then, after the class saw the results of the initial poll, she encouraged students to again discuss in small groups before polling a second time. In the first semester, she only polled students once and no opportunity for peer discussion followed the poll. Irene told us, “This is new for me this year. This was actually suggested to me by our [center for teaching and learning]. I really like the suggestion because it allows the students to think about their answer.” The distribution of students' responses shifted towards a majority selecting the correct answer after the second poll. Irene explained why she thought this shift had occurred, “If you just poll them [once], they're not having to explicitly explain why. You could just pick an answer, but talking to somebody, you actually have to say why… By that, you are having to engage with the material more to have that conversation, I think, and that can lead to a deeper understanding of it and just realizing that maybe, ‘Oh, my initial impression was wrong.’ It gives them time to re-evaluate it.” In addition to adding the idea that explaining reasoning fosters learning, Irene offered that the mechanism leading to learning is students realizing something about their initial thinking was wrong (i.e., generative work as a mirror).

Irene also modified the framing of questions she posed to students in ways that created more opportunities for generative work. In semester one, Irene posed a multiple-choice question on a slide that included two possible processes that could result in an observed pattern and an option to indicate that there was “not enough information to distinguish,” which was the most scientifically accurate choice (Figure S4.2). She then described an experiment that could be conducted to determine which process caused the observed pattern and displayed two graphs with data that could result from the experiment. The text on the slide asked students which graph corresponded with one of the processes and why. However, when Irene verbally posed the question during class, she did not ask for reasoning and told students “If you feel like it, feel free to put this in the chat, and then we’ll discuss the answer.” Students submitted answers but did not share reasoning behind their selection. In the second semester, Irene asked an open-ended question rather than a multiple-choice question about which process could drive the observed pattern and did not suggest that insufficient information had been provided (Figure S4.2). Students discussed with their neighbors and entered responses ranging from one word to a short sentence or two. Irene displayed these responses to the class and pointed out that some students had chosen each process, and that they were collectively correct to identify that both processes could explain the result.

Her next teaching move stood in contrast to the implementation of this activity in the prior semester. Whereas she had asked students in semester one to identify which of two plots represented a particular process, Irene challenged students in semester two to determine how to experimentally differentiate between the two processes before she showed hypothetical data. Irene then prompted students to "have a discussion." Reflecting on this open-ended question, Irene told us that " it's a good question because it's not the kind of question I've asked before very much. I've realized that this semester I am doing more of this experimental design… There wasn't just a single right answer… I felt they were understanding more than I have in the past. Were all the answers perfect? No, but I think getting them to think [about] how they would make that decision, it's a more creative process. They're having to engage in a higher-level learning ability, trying to critically think, analyze, and then create. I suppose that's what I'm trying to do is give those small opportunities to really get creative. Cause this is a large class. I can't give them projects that I can grade. I can't do short answer long answer kind of questions. So, it's like this is the opportunity for them to be a little more creative." Irene viewed the open-ended questions that she added as opportunities for generative work, which she described as "creative" in nature. Irene considered these opportunities valuable even if not all students ultimately arrived at a correct answer and considered them feasible even in her large class. Irene summed up her thinking on her new approach, "Learning is not just memorization. There are many other forms of learning. Being able to apply your knowledge in various ways is really important. [Students] have a deeper understanding of the subject if they can apply and be creative with that knowledge."

Irene's teaching practices changed in the second semester to include additional opportunities for students to discuss their reasoning with their peers after an initial poll. She also changed a closed-ended question to an open-ended question involving experimental design. These adjustments reflected her developing understanding that generative work and the process of engaging students in explaining their reasoning fosters deeper learning. We also gained insight into the influence of Irene's teaching professional development, as she attributed her decision to add a second opportunity for peer discussion to a formal training experience with her institution's center for teaching and learning. Irene's case demonstrates how pedagogical knowledge of how people learn can influence both nuanced and fundamental aspects of active-learning instruction.

Discussion

The central finding of this research is that college faculty who are committed to using active-learning strategies vary considerably in the teaching expertise they bring to this endeavor and their knowledge can influence how they implement teaching strategies. Nuanced differences in the implementation of active-learning strategies can impact students' learning experiences and outcomes (*e.g.,* Knight et al., 2013; Eddy et al., 2015; Knight et al., 2015). Effectively engaging students in generative work is not just about adopting a particular strategy, it involves designing tasks that require generative work, launching a task using a framing that promotes generative cognitive engagement, and maintaining the cognitive demand of the tasks during interactions with students (Tekkumru-Kisa et al., 2022). Our findings suggest that pedagogical knowledge about how people learn informs these teaching actions.

Knowing how people learn influences instructional design, launch, and implementation

Our participants demonstrated that pedagogical knowledge development corresponds with changes to instructional design, framing, and implementation likely to benefit student outcomes. In discussing the design of active-learning activities, participants often talked about generative work as a mirror. This pedagogical idea relates to the concept of metacognition, which is a student's awareness and control of thinking for learning (Cross and Paris, 1988; Stanton et al., 2021; Halmo et al., 2022). Metacognitive skills support students as they learn via self-explaining behaviors described at the generative levels of the ICAP framework (Chi & Wylie, 2014). Students with well-developed metacognitive skills tend to exhibit higher academic achievement and demonstrate more expert-like approaches to learning (Stanton et al., 2021). Several strategies (reflective questions, group work, polling and re-polling, etc.) exist for scaffolding students through the process of planning, monitoring, and evaluating their own learning (e.g., Tanner 2012; Stanton et al., 2021). Therefore, it is important for instructors to leverage their knowledge of generative work as a mirror when they design in-class activities.

Irene demonstrated knowledge of generative work as a mirror in the first semester post-lesson interview and in each subsequent interview. In the second semester pre-lesson interview, Irene incorporated her knowledge of the importance of generative work as a mirror as she changed the overall design of her lesson to offer students a second chance to discuss their thinking with peers following an initial poll, creating an opportunity for students to think metacognitively. Irene's instructional change created opportunities for students to engage in metacognition, and her rationale for this change drew on her knowledge of generative work as a mirror. Importantly, this suggests pedagogical knowledge can enable instructors to implement practices that create opportunities for metacognition, a widely studied, crucial component of student learning in the active-learning classroom (e.g., Tanner, 2012; Stanton et al., 2021; Halmo et al., 2022).

Participants also demonstrated how pedagogical knowledge influences teaching practices during the launch of active-learning activities. Participants exhibited two different rationales for group work: the benefit of hearing ideas from other students and the value of students explaining their reasoning to one another. These rationales are not mutually exclusive, but they could influence how instructors launch tasks. If instructors primarily see group work as a way for students to hear ideas from peers, they may use prompts such as, "Talk to your group about this question and see what they're thinking." Whereas an instructor prioritizing opportunities for students to explain their thinking might launch a task by saying, "Discuss your thinking with your group, and focus on the reasons for your answers. Then, I'll ask you to share your reasons" (Knight et al., 2013).

When instructors specifically direct students to use reasoning in their discussions with peers, their discussions are more likely to include reasoning connected to evidence, and their learning outcomes improve (Knight et al., 2013). Irene's case is one example of how pedagogical ideas about how people learn can impact the launch of a task. In the first semester, Irene's launch of a task did not prompt for reasoning and students did not share reasoning. In the second semester, she re-designed the question to be open-ended and asked [include prompt only if it's also included in the case]. Irene's change in practice coincided with her development of pedagogical knowledge of how people learn, specifically the idea that to explain is to learn, and would be predicted to benefit student learning outcomes based on prior research about the impact of instructional cues (Knight et al., 2013; Knight et al., 2015).

This study's methodological approach allowed us to access pedagogical knowledge instructors enacted during the implementation of active-learning activities. Participants seemed intuitively aware of one of the core principles of generative work in the ICAP framework. Dana stated this core principle in simple terms "You don't know something until you have to teach it." At some point in their experiences as learners, teachers, or both, several of our participants had developed this salient understanding of the effectiveness of generative work for learning. When examining the time instructors spent interacting one-on-one versus the time they spent hearing student reasoning, it became clear that a salient understanding of "learning by teaching" did not necessarily translate to an instructor's teaching practices. For instance, Dana exhibited her awareness that to explain is to learn in all four semesters of her participation. Until the final semester, she frequently undermined opportunities for generative work by walking students through logical progressions toward the answer instead of prioritizing space for hearing students' explanations.

Instructors commonly struggle to balance the need to engage students in generative work against the need for students to arrive at a canonical understanding. So, they tend to steer students toward an answer as they do much of the work in providing reasoning and engage in limited form of dialogue as they ask students to recall information (Kranzfelder et al., 2020). Dana was not alone in experiencing these difficulties, both in our sample and among instructors from other studies. Her improvements made over the course of the study set her apart from many active-learning instructors who undermine learning opportunities by being overly directive. Importantly, an instructor's ability to play the role of the "guide on the side" instead of the "sage on the stage" as they engage in interactive dialogue with students positively predicts learning outcomes (Hake et al., 1998; Deslauirers et al., 2011; Kranzfelder et al., 2020). Leveraging the already salient idea of "learning by teaching" present among instructors could help instructors adjust their approach, as Dana eventually did, to elicit student reasoning during generative work to better align the nuances of her teaching implementation with the ICAP framework.

This underscores the importance of integrating key insights from teacher knowledge research into faculty training programs and providing additional support for college instructors seeking to improve their teaching practices. Indeed, research on teaching knowledge can play a pivotal role in shaping teaching professional development opportunities (*e.g.*, Fennema et al., 1993; Hill et al., 2005). As such, efforts focused on developing active-learning instructional strategies and facilitating their adoption should be paired with research on the teaching knowledge that instructors need to effectively implement these practices. In particular, knowledge of how people learn can prepare instructors to adapt teaching strategies to their students and context without undermining the critical features that support student learning (Smith, 2015). Toward this end, findings from our study expand upon prior work (Auerbach et al., 2018; Auerbach & Andrews, 2018; Andrews et al., 2019) by identifying specific pedagogical ideas about how students learn and nuances of instructional practice change that corresponded with the development of these ideas.

Productively expanding on common pedagogical ideas about how people learn

Participants relied on a range of productive pedagogical ideas that can be leveraged in teaching professional development. Positioning faculty as learners, we considered what ideas these learners might bring into a teaching professional development setting and how those ideas might develop. We drew on the ICAP framework and our findings to create a list of pedagogical ideas that were common among our participants and ways to productively expand on these ideas to move toward more expert-like pedagogical knowledge (i.e., based in research and theory), which we hope can be a resource to both faculty in their own self-reflection and to educational developers (Figure 4.2). An instructor's goal should not be to distance themself from any of these common pedagogical ideas because they are likely to be productive for active learning. Instead, our findings suggest that certain participants expanded on these common pedagogical ideas in ways that brought their knowledge in closer alignment with the ICAP framework. Since these expansions held saliency with our participants, they are likely to also resonate with the larger population. Furthermore, their connections to more common pedagogical ideas offer a concrete route for pedagogical knowledge development.

Common pedagogical idea		Productive expansion
Active learning helps keep my students' attention during class	AND	this time is most impactful if students engage in applying knowledge to novel scenarios and explaining their reasoning.
Active learning should be difficult, involving more than just recalling facts	AND	when students are supported sufficiently, difficulty can be a sign that they are learning, rather than the opposite.
Students benefit from hearing how their peers explain concepts	AND	they learn even more from the chance to explain their own thinking.
When I teach content to students, I learn it more deeply	AND	students can similarly learn when they have to explain their thinking to others.
Active learning involves going beyond the given information and committing to a line of reasoning	AND	this supports learning by reflecting their thinking back to them, enabling them to realize what they know and don't know.
When I interact with students, I try not to give them the answer away	AND	students learn more when I can support them in explaining a concept to me than when I explain it to them.

Figure 4.2. Common pedagogical ideas and productive expansions on these ideas made by some instructors.

A missing link: pedagogical knowledge connecting evidence-based teaching practices to positive student outcomes

Implicit in the structure and culture of the higher education institution is the assumption that its instructors will learn how to teach effectively on their own. The observer will find this assumption laid bare in the lack of formal pedagogical training required of early-career undergraduate instructors (e.g., Schussler et al., 2015), the lack of formalized teaching mentorship programs available to faculty (e.g., Brickman et al., 2016), and the lack of formal opportunities for instructors to receive feedback on their instruction (e.g., Brickman et al., 2016). If student learning represents an important objective of the higher education institution, we should interrogate a system that spreads instructors thin across their myriad roles as teachers of industrial-sized classrooms, managers of competitive research laboratories, mentors of the next generation's scientists, and members of their communities.

Though in many ways inherent to this system, the assumption that instructors will naturally learn to teach effectively is rarely scrutinized. Our findings offer a potential explanation for why effective teaching does not necessarily result from teaching experience alone. Many undergraduate instructors lack the pedagogical knowledge essential to effective active-learning instruction, and, without sufficient support, they face challenges in developing this knowledge amidst their myriad responsibilities. This study joins a relatively small, but growing body of work that suggests a role for pedagogical knowledge research in providing a crucial resource that can enable instructors to carry out effective implementations of active learning that capitalize on the ICAP framework.

Our study emphasizes the challenge faced by the modern undergraduate STEM instructor. Unlike K12 teachers who receive formal pedagogical training, many faculty are not participating in teaching professional development at all, nor do they receive training or mentorship for their teaching responsibilities. Mandatory teaching professional development for graduate teaching assistants in the life sciences is often fewer than 10 hours and focused on policies and classroom management (Schussler et al., 2015). Once in a faculty position, instructors typically do not receive sufficient mentoring nor feedback on their teaching (e.g., Brickman et al., 2016). Thus, supporting faculty in developing knowledge about how people learn likely requires shifting the expectations and incentives for faculty to invest in developing as teachers, and evaluating what learning opportunities are needed to support pedagogical knowledge development among STEM faculty.

Though the faculty we studied tended to participate in teaching professional development, they did not necessarily develop more expert-like pedagogical knowledge during this study. As faculty are increasingly encouraged to adopt active-learning strategies, we must ask whether the existing systems can support necessary expertise development. Our data highlight the potential positive impact of institutional support for teaching, such as when Irene modified her instructional approach to invite students to discuss questions after they voted individually based on guidance for a local institutional center that provides teaching professional development. The guidance she received seemed to support both her knowledge development and a shift in instructional practices. In other cases, participants engaged in extensive teaching development opportunities and had yet to develop more expert-like knowledge of how people learn. This suggests that teaching professional development may need to be intentionally designed to help faculty develop this expertise.

Pedagogical knowledge about how people learn can equip instructors to adapt and fine-tune their teaching strategies without undermining the critical features that support student learning (Smith, 2015). For example, about 40% of faculty who use peer instruction do not consistently have students discuss ideas in groups (Dancy et al., 2016), yet this component positively impacts student learning (Vickrey et al., 2015). An instructor who relied on the pedagogical idea that students learn from explaining their thinking would not choose to skip this part of peer instruction because they would view it as the part of the strategy that most supported student learning. Studies examining the details of how instructors implement evidence-based strategies consistently find that critical components—those that make the strategy effective for student learning—are modified and undermined (e.g., Dancy et al., 2016).

This study suggests that pedagogical knowledge could be an essential link between evidence-based strategies and achieving the desired student outcomes. Emphasizing pedagogical knowledge about how people learn in professional development can help instructors stay aligned with the intended use of evidence-based practices, even as they adapt these practices to suit their specific teaching contexts. By the same token, paying close attention to the teaching knowledge required for effectively implementing evidence-based strategies could aid the developers of these strategies in expanding their successful application and ultimately enhancing student outcomes. However, testing student outcomes in association with teacher knowledge remains an important goal for future work. By understanding the relationship between teacher knowledge, teaching practices, and student outcomes, we might find teaching knowledge to be the essential missing link that connects evidence-based teaching practices to their hypothesized student benefits.

Future directions in teacher knowledge research

Teaching knowledge for effective active-learning instruction at the undergraduate level remains understudied, but recent work has established a foundation that future research can build on. Excluding the present study, only seven peer-reviewed papers have investigated teaching knowledge for evidence-based teaching in undergraduate STEM classrooms (Andrews et al., 2022; Chapter 3). This work has identified a framework of core components of pedagogical knowledge, including knowledge of how people learn, and how this knowledge varies between expert and novice instructors (Auerbach et al., 2018; Auerbach & Andrews, 2018). The present study builds on prior work by identifying additional variation within pedagogical knowledge of how people learn and characterizing the alignment of this knowledge to the knowledge predicted to be important by the ICAP framework. Critically, robust evidence on the relationship between pedagogical knowledge and student learning outcomes. While studies have explored teacher knowledge and teaching practices separately, as well as the link between teaching practices and student outcomes, there is still a need for research that examines how teaching knowledge, practices, and student outcomes interact. Such studies are essential to determine which aspects of pedagogical knowledge contribute to effective teaching and how pedagogical knowledge impacts which students benefit from active-learning instruction.

As our longitudinal study progressed, questions arose about participants' pedagogical knowledge which could serve as a focus for future research. Participants frequently referenced the idea of generative work as a mirror when explaining their rationales for using active-learning multiple-choice questions. Although we correctly assessed that these activities involved generative work, participant responses led us to wonder whether these participants focused primarily on metacognition that occurred only after the answer had been shown and not during the time allotted for student work. Relatedly, some participants thought about generative work as providing an opportunity for students to realize whether they were generally right or wrong whereas other instances of this knowledge focused on generative work as an opportunity for students to realize which aspects of the problem they had gotten wrong and why. Additionally, some participants exhibited knowledge of the importance of social and individual metacognition, some exhibited knowledge of one or the other, and a few exhibited knowledge of neither. This suggests participants vary in their understanding of which aspects of generative work as a mirror are important which could impact their decisions to use group work. We could not reliably assess any of these prospects qualitatively. Using more targeted interview questions, an in-depth characterization of teacher knowledge pertaining to student metacognition could provide valuable insights.

A key finding of this study is that participants often recognized the importance of generative work, but evidence of their understanding sometimes lacked a crucial component for effectively facilitating it: the importance of refraining from explaining during interactions with students. If instructors approach their interactions with students with a primary goal of avoiding giving away the answer, they could undermine learning opportunities by walking students through their own logical progression. They could do this and still achieve their goal so long as students demonstrate their ability to arrive at the correct answer by making the final leap of the logical progression. We saw Dana exhibit this kind of thinking and such teaching practices early in the study. We also saw a delay between when she developed knowledge of refraining from explaining and her ability to refrain from explaining in practice, suggesting we did not capture important details of her knowledge that enabled her to refrain from explaining in the final semester. This raises the possibility that certain knowledge allows instructors to view the goal of their interactions with students differently. Characterizing this knowledge directly could provide insights to support instructors in refraining from explaining as they interact with students during class.

Limitations

We urge readers to carefully consider the study's limitations before drawing conclusions from its findings. First, our sample of participants differs from the typical undergraduate instructor population in several important ways, limiting the generalizability of our findings to all biology faculty. Notably, many of our participants had a high level of engagement in teaching professional development, with the majority having accumulated over 40 hours of such training and a subset having experience in education research. This level of commitment suggests that these instructors were particularly invested in improving their teaching and likely had more opportunities to develop pedagogical knowledge than the average faculty member.

Another potential limitation is the impact of the COVID-19 pandemic on our longitudinal data collection. The pandemic forced many instructors to adapt their target lessons to fit new instructional modalities, which likely influenced the pedagogical knowledge and active-learning strategies they were able to demonstrate. For instance, remote instruction makes it more difficult to facilitate group work during class which affected several participants' active learning strategies, which affected the discussions we had during interviews regarding their instructional design and implementation. However, we took steps to mitigate these impacts during interviews by prompting participants to avoid discussions of COVID-related instructional changes, and we gently guided them away from these discussions when they emerged. Additionally, while teaching adaptations during the pandemic undoubtedly influenced the manifestations of participants' knowledge during interviews, we regarded participants' fundamental understanding of how people learn as unlikely to change in ways that would prohibit the use of these data. We carefully considered how these shifts in modality may have affected our ability to elicit and discern pedagogical knowledge, and we did not detect patterns that would invalidate our findings (Figure S4.1). However, the unique context of the pandemic should be considered when interpreting the results.

By engaging in the interviews, we likely solicited reflection on teaching that participants would not have otherwise engaged in. This could have led to knowledge and practice development that might not have otherwise occurred. This limitation is difficult to avoid in longitudinal qualitative studies. However, one take-away of this study for instructors is the value of reflecting on why you are using a particular active-learning strategy or activity. Participants sometimes commented during interviews that they had not had previous opportunities to reflect on their teaching as deeply as they did during interviews for this study. We believe this reflects a common reality faced by faculty; it requires a lot of intentionality to make time to reflect on one's teaching because it’s not an expectation of our positions. Yet, reflecting on the purpose of a particular strategy can help an instructor recognize implicit goals and underlying assumptions that may be in play and point toward priorities for implementing the strategy to ensure it meets the intended goals.

Finally, our study focused on a single focal lesson to characterize instructors’ pedagogical knowledge. This approach does not attempt to fully represent their overall teaching practices. If we had observed a different lesson, our findings would likely differ. For the purposes of our longitudinal design, this focus was key to observing pedagogical knowledge development and changes to lesson design, framing, and implementation over time.

Conclusion

Instructors' teaching knowledge plays a crucial role in the successful implementation of evidence-based practices in STEM education. While much research on teaching practices emphasizes the strategies themselves, it often overlooks the foundational knowledge instructors need to use these practices effectively. By focusing on the development of teaching knowledge, we can enhance the adoption and impact of these strategies. Developers and researchers should prioritize teaching knowledge from the outset, ensuring that evidence of teaching strategy effectiveness includes the necessary resources and knowledge for instructors. Ultimately, the success of evidence-based practices in improving student outcomes relies on their thoughtful and informed implementation by faculty and advancing STEM education will require directly addressing the teaching knowledge faculty need while also rewarding those who pursue this type of professional growth.

References

Alonzo, A. C., & Kim, J. (2016). Declarative and dynamic pedagogical content knowledge as elicited through two video-based interview methods. *Journal of Research in Science Teaching*, *53*(8), 1259-1286.

Andrews, T. C., Auerbach, A. J. J., & Grant, E. F. (2019). Exploring the relationship between teacher knowledge and active-learning implementation in large college biology courses. *CBE—Life Sciences Education*, *18*(4), ar48.

Andrews, T. M., Leonard, M. J., Colgrove, C. A., & Kalinowski, S. T. (2011). Active learning not associated with student learning in a random sample of college biology courses. *CBE—Life Sciences Education*, *10*(4), 394-405.

Andrews, T. C., Speer, N. M., & Shultz, G. V. (2022). Building bridges: a review and synthesis of research on teaching knowledge for undergraduate instruction in science, engineering, and mathematics. *International journal of STEM education*, *9*(1), 66.

Anfara Jr, V. A., Brown, K. M., & Mangione, T. L. (2002). Qualitative analysis on stage: Making the research process more public. *Educational researcher*, *31*(7), 28-38.

Auerbach, A. J. J., & Andrews, T. C. (2018). Pedagogical knowledge for active-learning instruction in large undergraduate biology courses: a large-scale qualitative investigation of instructor thinking. *International journal of STEM education*, *5*, 1-25.

Auerbach, A. J., Higgins, M., Brickman, P., & Andrews, T. C. (2018). Teacher knowledge for active-learning instruction: Expert–novice comparison reveals differences. *CBE—Life Sciences Education*, *17*(1), ar12.

Ballen, C. J., Wieman, C., Salehi, S., Searle, J. B., & Zamudio, K. R. (2017). Enhancing diversity in undergraduate science: Self-efficacy drives performance gains with active learning. *CBE—Life Sciences Education*, *16*(4), ar56.

Brickman, P., Gormally, C., & Martella, A. M. (2016). Making the grade: Using instructional feedback and evaluation to inspire evidence-based teaching. *CBE—Life Sciences Education*, *15*(4), ar75.

Chi, M. T. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in cognitive science*, *1*(1), 73-105.

Chi, M. T., Adams, J., Bogusch, E. B., Bruchok, C., Kang, S., Lancaster, M., ... & Yaghmourian, D. L. (2018). Translating the ICAP theory of cognitive engagement into practice. *Cognitive science*, *42*(6), 1777-1832.

Chi, M. T. (2018). Learning from examples via self-explanations. In *Knowing, learning, and instruction* (pp. 251-282). Routledge.

Chi, M. T., Kang, S., & Yaghmourian, D. L. (2017). Why students learn more from dialogue-than monologue-videos: Analyses of peer interactions. *Journal of the Learning Sciences*, *26*(1), 10-50.

Chi, M. T., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational psychologist*, *49*(4), 219-243.

Cross, D. R., & Paris, S. G. (1988). Developmental and instructional analyses of children's metacognition and reading comprehension. *Journal of educational psychology*, *80*(2), 131.

Dancy, M., Henderson, C., & Turpen, C. (2016). How faculty learn about and implement research-based instructional strategies: The case of peer instruction. *Physical Review Physics Education Research*, *12*(1), 010110.

Deslauriers, L., Schelew, E., & Wieman, C. (2011). Improved learning in a large-enrollment physics class. *science*, *332*(6031), 862-864.

Eddy, S. L., & Hogan, K. A. (2014). Getting under the hood: How and for whom does increasing course structure work?. *CBE—Life Sciences Education*, *13*(3), 453-468.

Eddy, S. L., & Hogan, K. A. (2014). Getting under the hood: How and for whom does increasing course structure work?. *CBE—Life Sciences Education*, *13*(3), 453-468.

Eddy, S. L., Converse, M., & Wenderoth, M. P. (2015). PORTAAL: A classroom observation tool assessing evidence-based teaching practices for active learning in large science, technology, engineering, and mathematics classes. *CBE—Life Sciences Education*, *14*(2), ar23.

Fennema, E., Franke, M., Carpenter, T., & Carey, D. (1993). Using children's mathematical knowledge in instruction. American Educational Research Journal, 30(3), 555–583.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the national academy of sciences*, *111*(23), 8410-8415.

Guba, E. G. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries. *Ectj*, *29*(2), 75-91.

Haak, D. C., HilleRisLambers, J., Pitre, E., & Freeman, S. (2011). Increased structure and active learning reduce the achievement gap in introductory biology. *Science*, *332*(6034), 1213-1216.

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American journal of Physics*, *66*(1), 64-74.

Halmo, S. M., Bremers, E. K., Fuller, S., & Stanton, J. D. (2022). "Oh, that makes sense": Social Metacognition in Small-Group Problem Solving. *CBE—Life Sciences Education*, *21*(3), ar58.

Hill, H., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. American Educational Research Journal, 42(2), 371–406.

Jackson, M. A., Moon, S., Doherty, J. H., & Wenderoth, M. P. (2022). Which evidence-based teaching practices change over time? Results from a university-wide STEM faculty development program. *International Journal of STEM Education*, *9*(1), 22.

Knight, J. K., Wise, S. B., Rentsch, J., & Furtak, E. M. (2015). Cues matter: Learning assistants influence introductory biology student interactions during clicker-question discussions. *CBE—Life Sciences Education*, *14*(4), ar41.

Knight, J. K., Wise, S. B., & Southard, K. M. (2013). Understanding clicker discussions: Student reasoning and the impact of instructional cues. *CBE—Life Sciences Education*, *12*(4), 645-654.

Kranzfelder, P., Bankers-Fulbright, J. L., García-Ojeda, M. E., Melloy, M., Mohammed, S., & Warfa, A. R. M. (2020). Undergraduate biology instructors still use mostly teacher-centered discourse even when teaching with active learning strategies. *BioScience*, *70*(10), 901-913.

Menekse, M., & Chi, M. T. (2019). The role of collaborative interactions versus individual construction on students' learning of engineering concepts. *European Journal of Engineering Education*, *44*(5), 702-725.

Menekse, M., Stump, G. S., Krause, S., & Chi, M. T. (2013). Differentiated overt learning activities for effective instruction in engineering classrooms. *Journal of Engineering Education*, *102*(3), 346-374.

Nehm, R. H., Finch, S. J., & Sbeglia, G. C. (2022). Is active learning enough? The contributions of misconception-focused instruction and active-learning dosage on student learning of evolution. *BioScience*, *72*(11), 1105-1117.

Offerdahl, E. G., McConnell, M., & Boyer, J. (2018). Can I have your recipe? Using a fidelity of implementation (FOI) framework to identify the key ingredients of formative assessment for learning. *CBE—Life Sciences Education*, *17*(4), es16.

Pollock, S. J., & Finkelstein, N. D. (2008). Sustaining educational reforms in introductory physics. *Physical Review Special Topics—Physics Education Research*, *4*(1), 010110.

Schussler, E. E., Read, Q., Marbach-Ad, G., Miller, K., & Ferzli, M. (2015). Preparing biology graduate teaching assistants for their roles as instructors: An assessment of institutional approaches. *CBE—Life Sciences Education*, *14*(3), ar31.

Shenton, A. K. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for information*, *22*(2), 63-75.

Smith, G. A. (2015). Why college faculty need to know the research about learning. *InSight: A Journal of Scholarly Teaching*, *10*, 9-18.

Smith, M. K., Jones, F. H., Gilbert, S. L., & Wieman, C. E. (2013). The Classroom Observation Protocol for Undergraduate STEM (COPUS): A new instrument to characterize university STEM classroom practices. *CBE—Life Sciences Education*, *12*(4), 618-627.

Stains, M., & Vickrey, T. (2017). Fidelity of implementation: An overlooked yet critical construct to establish effectiveness of evidence-based instructional practices. *CBE—Life Sciences Education*, *16*(1), rm1.

Stanton, J. D., Sebesta, A. J., & Dunlosky, J. (2021). Fostering metacognition to support student learning and performance. *CBE—Life Sciences Education*, *20*(2), fe3.

Tanner, K. D. (2012). Promoting student metacognition. *CBE—Life Sciences Education*, *11*(2), 113-120.

Tekkumru-Kisa, M., Coker, R., & Atabas, S. (2022). Learning to teach for promoting student thinking in science classrooms. *Teaching and Teacher Education*, *120*, 103869.

Theobald, E. J., Hill, M. J., Tran, E., Agrawal, S., Arroyo, E. N., Behling, S., ... & Freeman, S. (2020). Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math. *Proceedings of the National Academy of Sciences*, *117*(12), 6476-6483.

Turpen, C., & Finkelstein, N. D. (2009). Not all interactive engagement is the same: variations in physics professors' implementation of peer instruction. *Physical Review Special Topics—Physics Education Research*, *5*(2), 020101.

Vickrey, T., Rosploch, K., & Stains, M. (2015). Measurement of Faculty's Fidelity of Implementation of Peer Instruction following an Intensive Professional Development Workshop.

Supplemental Materials

Table S4.1. Knowledge quantification values based on presence and absence of pedagogical ideas about how people learn. This quantification was applied separately to pre- and post-lesson interviews. Possible total values could range from 2 to 10. The presence of refraining from explaining in an interview earned participants an extra bonus point independent of the other codes present. Other possible code combinations exist, but quantified value depend only on the presence of the codes listed.

Code combination	**Quantification Value**
Active rather than passive and/or Peer-to-peer listening and/or Struggle makes it stick	1
Application and practice	2
To explain is to learn	2
Application and practice and Generative work as a mirror	3
To explain is to learn and Generative work as a mirror	3
To explain is to learn and Application and practice	3
To explain is to learn and Application and practice and Generative work as a mirror	4
Refraining from explaining	+1 (bonus)

Table S4.2. Pedagogical knowledge of how people learn codebook. Code names and descriptions are provided. Two ideas, however, stood out from the rest. One pedagogical idea, which we termed “generative work as a mirror,” encompassed an idea about the mechanism by which generative work fosters student learning. Another pedagogical idea, which we refer to as “refraining from explaining” emerged when instructors described their intentions as they facilitated student work in real-time during class. It is critical that these segments show the instructor's thinking, ideas, and rationales, not just their practice.

Code Name	Description
Active rather than passive	Captures instances where an instructor explains that they are trying to move beyond traditional lecture in relation to cognitive engagement. Mostly commonly, these quotes discuss trying to go beyond/replace/avoid recall, memorization, algorithmic problem-solving, or lecture. They may talk about breaking up lectures. These ideas tend to be vague and may largely reflect a desire to move beyond passive cognitive engagement. These ideas represent the potential for a shift in thinking about what sort of student and teacher activity promotes learning.
Peer-to-peer learning	Captures evidence that instructors think that students learn from getting the chance to hear ideas or a diversity of ideas in groups. The instructor might state that this improves student learning explicitly or they might discuss something more vague. These sometimes vaguely refer to feedback but it’s not clear if the students get feedback just by listening to the ideas of others or if they directly provide each other with feedback. It’s not clear from these quotes if the instructor is thinking about anything beyond listening and maybe talking. Sometimes, the instructor emphasizes the “strength in numbers” advantage of group work in which each group has an increased likelihood of arriving at a correct answer compared to any individual student.
Struggle makes it stick	Captures evidence that the instructor thinks that a task helps students learn because students get it wrong or struggle, but the instructor offers limited or no reasoning about why mistakes and struggle are helpful. The instructor may explain that struggling helps students because they will remember better things they got wrong, or they will pay more attention after they get something wrong. Instructor may contrast struggling with just listening or writing things down, or contrast struggling with seeing a problem solved. They may also discuss the importance of students getting an answer wrong so that they are primed to learn about it. If an instructor’s rationale is that struggle helps students

	because they realize what they do and do not understand, that should be coded as generative work as a mirror instead.
Application and practice	Captures evidence that the instructor's rationale for using tasks (questions, problems, cases, etc.) that ask students to apply concepts or definitions, practice using concepts and definitions, using or doing something with some particular knowledge or understanding. They may describe this as applying to new scenarios, but a "new scenario" is not necessary to earn this code. These quotes generally align with students doing generative work. Use this code only for instances when the instructor does not clearly describe the students having to explain their understanding in some way (explaining, drawing, creating a model). This code needs to include some thought towards what students will have to do. Participants need to go beyond just naming the task.
To explain is to learn	Captures evidence that the instructor's rationale for using tasks (questions, problems, cases, etc.) that ask students to explain, externalize, or represent their understanding (and maybe reasoning) in some way to the outside world. This may involve students explaining their thinking verbally or in writing, drawing, etc. This would include interactive generative work (no separate code because it is so rare). These quotes align with students doing generative work. This excludes things like brainstorming ideas or making guesses. Answering a clicker question or other close-ended question is not sufficient. It isn't necessary that the instructor is there to see or hear the explanations students are giving. This code needs to include some thought towards what students will have to do. Participants need to go beyond just naming the task.

Generative work as a mirror	Captures evidence that the instructor thinks that the generative work that they ask students to do allows the students to recognize what they do and/or do not know. This might include students realizing that they have a misconception. The instructors think that students will have these realizations when they have to complete a task and cannot quite do it or get it wrong. These quotes reveal what the instructor thinks about how students learn but they may or may not explicitly refer to learning, mastery, understanding. This code includes either/or intentions for students to learn in-the-moment as well as prompting them to study certain things after the lesson. These ideas look different when discussed in the context of clicker questions that lack discussion.
Refraining from explaining	Captures evidence that the instructor is deliberately and strategically taking action with the goal of students doing generative work. These instructors are aiming to prompt students to be generative, such as asking students to provide a rationale, explanation, or to elaborate on their thinking. This knowledge relates to instances where the instructor is talking to small groups or the whole class, while they are working in class, with the goal of helping students be responsible for doing the cognitive work. Their practices may include deliberately not giving students answers, answering a question with a question, walking away before students have the full answer, specifically asking for reasoning, showing restraint in the face of inaccurate student answers. In an instance of this knowledge, the instructor is not aiming to be the main source of information when facilitating group work. These actions are more likely to be dynamic, in-the-moment decisions made by the instructor in response to students' reasoning or other evidence of their cognitive engagement.

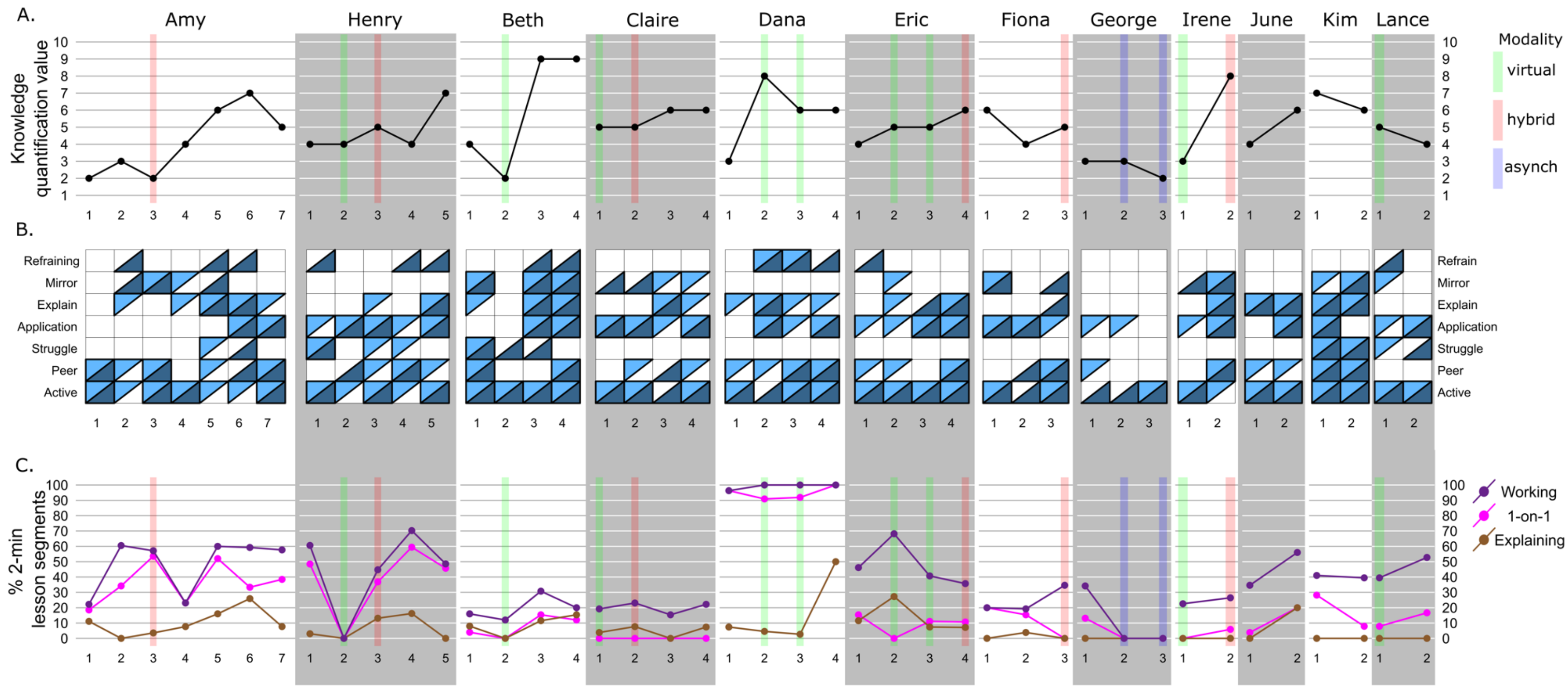


Figure S4.1. Longitudinal trajectories of pedagogical knowledge of how people learn, by participant. (A) Same data shown in Figure 4.1A. Hybrid lessons were held with some students in the classroom and others joining remotely via teleconferencing software. Virtual lessons involved a full-online, synchronous lesson given via teleconferencing software. Asynchronous lessons involved pre-recorded video lectures. Synchronous, in-person lessons are denoted by the absence of a colored vertical bar. (B) Same data shown in Figure 4.1B, but with triangles to indicate whether the subcode emerged in the pre- (light blue), post-lesson (dark blue) interview, or both. (C) Observation protocol data showing the percentage of two-minute lesson segments in which the instructor interacted one-on-one with students (pink), students were working (purple), or students were explaining their reasoning (brown). During data collection, the COVID-19 pandemic forced many participants to change the modality of their focal lesson. Modality of each focal lesson is denoted with colored vertical bars (A & C): hybrid (red), asynchronous (blue), and virtual (green).

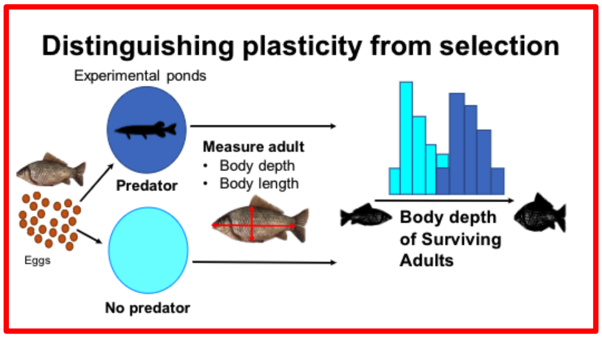
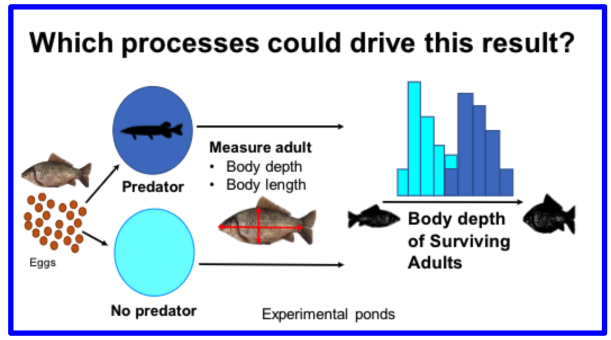
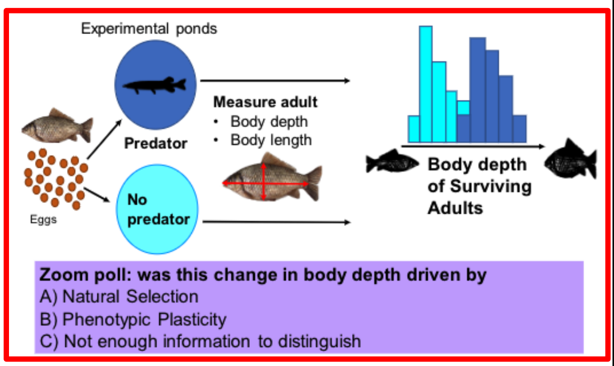
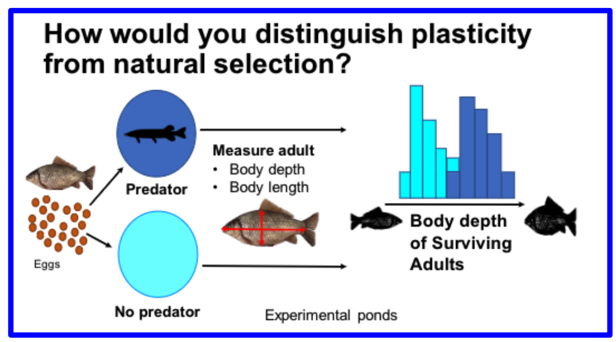
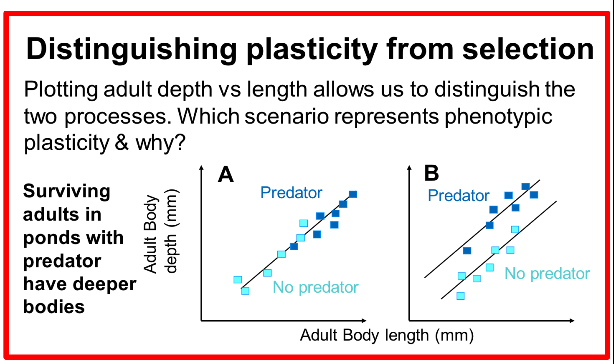
Semester 1	Semester 2
	
	
	

Figure S4.2. Irene's slides used during class to teach the distinction between plasticity and natural selection. Slides from semester 1 (red outline) and semester 2 (blue outline) are shown.

Table S4.3. Interactions between Dana and her students pertaining to the same question in each focal lesson. Lesson transcripts from each semester (Sem.) are provided.

Sem.	Dana's Lesson Transcripts
1	Student A: [referring to question 7] We didn't really understand the differentiation? Is it talking about secondary? **Dana: So, differentiation of the gonad to testosterone-producing tissue, that's basically differentiating into testes, right? So, that's happening during primary sex determination, right? And then differentiating into estrogen-producing, well that's going to be the response of beta-catenin and basically promoting that ovary fate, right? So you have an XY individual where we're making SOX9, so the job of SOX9 is to promote testes fate, and normally it will repress female fate by repressing beta-catenin, right? If it can't repress beta-catenin, does beta-catenin get expressed?** Student B: Yes **Dana: Yes, normally it gets expressed until it gets repressed by SOX9. So, for female, well, there's no Y gene, no SRY, no SOX9. That's why beta-catenin takes over and helps make sure there's not going to be any of that by repressing that fate and then promotes ovary fate. So, in a male, there's still beta-catenin sort of at some level, but once SOX9 gets expressed, it shuts that down and says, "No, no, no, don't do any ovary fate. We are all testes fate here."** Student A: So, that's why we thought it would do both. **Dana: Okay. Yeah. Right. Yeah. Because there's nothing to repress that beta-catenin signal. Okay, so now you have both. So, the testes tissue, what hormones is it going to try and make?** Student B: Testosterone **Dana: Testosterone. And?** Student B: Anti-mullerian hormone. **Dana: Yeah. So, you do get anti-mullerian hormones expressed and you have testosterone. [points to correct answers which students had already selected] So, yeah. You're good.**
2	Student A: And then number seven? **Dana: All right. So, SOX9 and beta-catenin not being able to repress each other… are both expressed? Neither expressed? Only one expressed? What's happening?** Student A: Would both be expressed?

	Dana: Yeah, so both of them are going to be expressed. What is the effect of expressing Sox nine? Student A: Um, you have anti-mullerian hormone being expressed. **Dana: Yeah. What happens in terms of gonad development?** Student A: You form testes. **Dana: Okay. If beta-catenin is expressed, what happens in terms of gonadal development?** Student A: You form ovaries. **Dana: So now, if both are being expressed, what do you think is going to happen?** Student A: You would get both? **Dana: Yeah. Or at least try to get both. So, in terms of exactly what's happening here, it's going to be harder to figure out sort of with a thought experiment in that SOX9 not only promotes the testes formation, but represses ovary formation. Beta-catenin promotes ovary formation, but represses the testes formation. So, you're likely to end up with a gonad with some mixture. So, if you notice, there's wording of "differentiation of the gonad to testosterone producing tissue". It's not really forming a testis that would be typical, but it's going to try to do that. So, you're going to end up with a gonad that is some mix of both. So, then hormone wise, do we have testosterone present? Do we have estradiol present?** Student A: You would have both. **Dana: Yeah. So, you're going to have tissue producing high levels of both of them. So, the endocrine system is going to be a mess in this individual, but when we think about the persistence of the Wolffian duct, if testosterone is present [and] estradiol is present. What happens to that Wolffian duct?** Student A: [thinks silently for 15 seconds] **Dana: Or how about this? What happens to the Wolffian duct if there aren't any hormones present?** Student A: It wouldn't be there. **Dana: Okay. So, apoptosis without any hormones. What prevents it from a apoptosing?** Student A: Testosterone. **Dana: Okay. Do we have testosterone present?**

	Student A: Yes. **Dana: Yes. So, it's not going to apoptose. So there's your persistence that, yes, it will persist because there's going to be high levels of testosterone being produced by at least part of the gonad. Yeah. So all of them are true. Does that make sense?** Student A: Yes.
3	**Dana: [joining virtual breakout room that requested help using hand raise function] How can I help?** Student A: We're having trouble with question seven. So, if SOX9 and beta-catenin cannot repress each other, we're probably still going to have AMH expression. **Dana: So can I suggest, before you jump to the hormones, what is happening in primary sex determination? Because you got to figure out what gonads are present before you can figure out what hormones they will produce, right? So SOX9 and beta-catenin, are we affecting primary sex determination or secondary sex determination?** Student B: Primary. **Dana: Yeah. Yeah. You're good. So talk me through your logic on what's going to happen in primary sex determination in this individual.** Student A: Okay. So we have SRY, you know, so we're getting expression of SOX9. We also have—beta-catenin is always present at varying levels, depending on what we're talking about. **Dana: Yeah.** Student A: But it's XY, you know, it's like kind of male already. But they can't repress each other. **Dana: Okay. So is SOX9 at a high level of expression?** Student A: Yes. **Dana: Yes. [noticing student B confused facial expression] I couldn't quite tell if those were like, 'I'm not following' eyes or not.** Student B: [nods to confirm they are following] **Dana: Okay. Good. Thanks. And then is beta-catenin expressed at a high level?** Student A: Yeah. **Dana: Yes. So both are there. So they're both going to be able to promote the fate they usually promote. They can't repress the other fate in this situation**

	because beta-catenin usually represses the male fate by repressing SOX9. SOX9 usually represses a female fate by repressing beta-catenin. If we take that away, now it's basically we can only promote the fates. We can't turn off the other fates. So what's going to happen in terms of the gonads? Student B: Both will be produced. **Dana: Exactly. And if you notice, I was a little careful with my wording that we're making estradiol-producing tissue. It's going to be a whole mix of a gonad. It is not going to look like a testes. It's not going to look like an ovary or it's going to be some combination, whether it's sort of here's a chunk of ovary-like and here's a chunk of testes-like, or it's kind of everything is everything. So you are going to make both a gonad that can make estradiol and testosterone. So you're going to have some testes-like tissue, some ovary-like tissue. So now we can start to think about the expression of hormones from that. Now that we know what the gonad is, now we can start to get into the hormones. So if we have both ovary-like tissue and testes-like tissue, what hormones are being produced?** Student B: Testosterone, estradiol, all those. Is dihydrotestosterone produced too? **Dana: Yep. The testes-like gonad is going to be trying its best to make testosterone, dihydrotestosterone, that's what it does. And then that ovary-like tissue is going to be trying its best to make estradiol. So, it's going to be a mess of hormones, but there's going to be all these hormones. But there's going to be all of it. So what about anti-mullerian hormone?** Student A: It's in the testes pathway. **Dana: Yeah. So that testes-like tissue also making anti-mullerian hormone in addition to testosterone. So you're right, that anti-mullerian hormone expression is definitely happening. What about the Wolfian duct?** Student B: Yeah. **Dana: If you have testosterone, it's going to stick around and differentiate. There's also anti-mullerian hormone expression. So the mullerian duct will go away, even though the estradiol is there. It's all about whether anti-mullerian hormone is present. In terms of external genitalia, I did not ask because I don't know either. It's going to be a relative level. And it's really hard to tell just from this, what's the relative amount of testosterone and estradiol. So that's why I didn't even get into that. It's like, which one's higher? I don't know, but both are going to be turned on. Does that make sense? [gesturing to question #7] So all of them are true on this one.**
4	Student A: Could you explain, for number seven, because they're not repressing each other, are both going to still take place? **Dana: I don't know. If the testes is not actively repressing ovary fate, can you get a mixed fate?**

	Student A: The simplest answer is the pathway... Whatever the pathway is would just continue. So, yeah. Right? **Dana: Say that again.** Student A: Like, if they don't repress each other, the simplest answer would be that they would just continue. **Dana: So, are you saying that you can get both ovary-like tissue and the testes-like tissue if they're not repressing each other?** Student A: Uhh Student B: I mean, if there's AMH expression, right? That would get rid of the female tissue. **Dana: Is AMH affecting primary sex determination?** Students A & B: No. **Dana: Okay. So, think about primary sex determination first, and then you can think about secondary and the hormones produced. So, SOX9 and beta-catenin can't repress each other. So then, are you going to get ovary-like tissue or testes-like tissue or both?** Student B: You would get both, right? **Dana: Why?** Student B: Because they're not repressing each other. **Dana: Yeah. So, and you'll note, I did not say we're forming a testes or an ovary full out. It's going to be some mixture of those tissues. So then, if you have those tissues, now think about what happens in terms of hormone production.** Student A: If you have both tissues, they're going to produce both hormones. But, like, even in a normal person, you're still going to produce testosterone and estradiol. **Dana: Right.** Student B: But they would just be more similar to, I guess, normal levels of both? **Dana: Right. Think about the levels that, you're right, that testosterone is going to be produced in XX individuals with no change. In typical females, just like estrogen is produced in typical XY individuals. But think about the relative level there. That it's not just, is it there or is it not, but how much? So, if you have testosterone-producing tissue, so testes-like tissue, how does that level compare to, like, just an ovary? How much testosterone do you think is going to be present compared to a typical female?**

	Student A: Probably higher, right? **Dana: Yeah. Okay. So it's going to be closer to a male-like level, or at least potentially. That's going to cause lots of estradiol to be produced. So basically, like, hormones are going to be a mess in this individual. But, you told me that there's going to be, like, relatively high levels of testosterone. How does that affect, in this case, we're thinking about the Wolfian duct?** Student A: The Wolfian duct degenerates as a result of increased expression of estradiol, right? **Dana: It's actually, it doesn't have to do with estradiol. It's going to be the presence or absence of testosterone. Yeah. So, since testosterone is there, it would still be there. Yep. It's going to persist. Because you told me that there's going to be, you know, relatively high levels, high enough levels of testosterone that that is likely to occur.** Student A: So does this individual just have everything? **Dana: It's going to be a mess. Yeah.**

CHAPTER FIVE

OPPPORTUNITIES FOR FUTURE RESEARCH

Overview

In this concluding chapter, I explore opportunities for future research to build upon the bodies of work presented in this dissertation. I begin by focusing on possible follow-up experiments to more robustly test the hypothesis of supergene-mediated genetic assimilation in fire ants. I also briefly outline the prospect of studying supergene-mediated genetic assimilation in other organisms. Next, I propose ideas for additional work to follow-up on my teacher knowledge research focused on interplay between pedagogical content knowledge and pedagogical knowledge of how people learn. This chapter ends with concluding remarks regarding this dissertation's focus on the "plasticity" of fire ants and human teachers.

Testing supergene-mediated genetic assimilation in fire ants

As discussed in the limitations in Chapter 2, our experimental design could not identify a causal link between the fire ant supergene and the increased genetic regulation of gyne weight and colony-founding behavior. However, our results and several aspects of fire ant biology and evolutionary history point to a hypothesis of genetic assimilation that needs further testing. A different experiment could more robustly test this hypothesis.

Levis and Pfennig (2016) provide a two-phase scheme for testing hypotheses related to plasticity-first evolution, a broad category of evolutionary phenomena that includes genetic assimilation. In the first phase, researchers make inferences to identify ancestral proxy lineages and a lineage displaying a canalized novel phenotype. Then, they can use a common garden design to construct reaction norms for the trait of interest across ancestral and novel environments. In fire ants, this would involve testing the reaction norm of monogyne *SB/SB,* polygyne *SB/SB,* and *Sb*-carrying polygyne gynes in a controlled environment that manipulates the inducing environmental variable(s) associated with overwintering. Reaction norms that show plasticity in the ancestral proxy lineages *(SB/SB*) and canalization of the phenotype in the derived lineage (*Sb*-carriers) would provide evidence consistent with genetic assimilation. Gene expression data collected to characterize these reaction norms could provide a more reliable set of genes associated with the possible supergene-mediated genetic assimilation in the fire ant system.

However, in fire ants, several challenges arise with this experimental design due. First, the precise environmental variable that induces overwintering plasticity is not known. It likely involves temperature, humidity, nestmate interactions, food resource availability, light-dark photoperiod, or some combination of these or other factors. Without the precise knowledge of the critical environmental variable(s), we cannot to design a common garden experiment that reliably isolates relevant variables. Second, polygyne *SB/SB* gynes are frequently attacked and executed by their nestmates during development. One study found a ratio of 30:1 between polygyne *SB/Sb* and *SB/SB* gynes at 11-14 days post-eclosion (Keller & Ross, 1999). Practically, this means any reaction norm comparison would likely involve monogyne *SB/SB* gynes in monogyne environments compared to *Sb*-carrying gynes in polygyne environments, a less-than-ideal setup that does not control for social environment (i.e., worker interactions). However, a common garden could be possible if the environmental variable could be identified and the problem of polygyne *SB/SB* addressed with large sample sizes or a clever experimental design.

Instead, it may be more tractable for future work to use different approaches and more robust sampling to establish a more reliable set of candidate genes associated with gyne nutritional plasticity as well as the supergene-mediated weight polymorphism. To do so, researchers could collect large samples of overwintering monogyne *SB/SB* gynes over a series of stages throughout the overwintering process. Sampling fat body tissue from each gyne for RNA sequencing, while also measuring gaster dry weight to quantify variation in fat body content could allow researchers to conduct a transcriptome-wide association study (TWAS) to identify dry-weight-associated SNPs and perform expression quantitative trait loci (eQTL) mapping to correlate gene expression levels with gaster dry weight.

With a robust set of plasticity-associated genes in hand, researchers could turn to investigate the effects of the supergene to create a comparable data set. By studying gene expression over the developmental period in which weight variation accumulates across *SB/SB* and *Sb*-carrying gynes, researchers could identify the candidate genes that *Sb* influences in the regulation of the weight polymorphism. Knowledge of the gyne weight polymorphism is largely limited to later stages of their development. The few studies of gyne weight during early adult development lacked molecular genetic data (Keller & Ross, 1999, DeHeer, 2002) and were conducted prior to the discovery of *Gp-9* or its association with the supergene (Keller & Ross, 1993a, 1993b). We have some knowledge of how genetics impacts this period of adult development from gene expression analyses of 1d post-eclosion adults, and 11d post-eclosion adults, but these studies lacked weight measurements (Wang et al., 2013; Nipitwattanaphon et al., 2013). Currently, we lack a detailed accounting of the weight variation that accumulates between *SB/SB* and *Sb*-carrying gynes during early adult development and the associated changes to gene expression. Collecting a similar set of RNA-sequencing data for TWAS and eQTL would yield a comparable set of candidate genes associated with the supergene's influence on gyne dry-weight.

Leveraging the recently completed PacBio genome assembly—which includes the full *Sb* region—one can identify plasticity-associated candidate genes with *Sb*-alleles that harbor non-synonymous mutations within *Sb*. Given that the *Sb* supergene influences the transcriptome both in *cis* and *trans* (Arsenault & King, et al., 2020), one could also investigate eQTL hits for their proximity to genes differentially expressed according to supergene genotype.

While this approach necessitates extensive sampling, it offers a more robust means of identifying coding and non-coding variants associated with variation in the overwintering phenotype, providing deeper insights into the genetic architecture of phenotypic plasticity in fire ant gynes. If candidate genes identified in the TWAS and/or eQTL mapping are disrupted by non-synonymous substitutions in the *Sb* region and/or exhibit differential expression by supergene genotype, the hypothesis of supergene-mediated genetic assimilation would find more robust support.

<u>Supergene-mediated genetic assimilation in other species?</u>

In addition to *S. invicta*, polygyny has been discovered to have genetic underpinnings in the form of supergenes in four other socially polymorphic ant lineages (Rosset & Chapuisat 2007; Avril et al., 2019; Blacher et al., 2021; Fontcuberta et al., 2021; Kay et al., 2022; Chapuisat et al., 2023). In these species, monogyny and polygyny are also regulated by a supergene, and the supergene regulates gyne weight and colony-founding polymorphisms akin to that of *S. invicta*. Their overwintering biology has yet to be characterized. However, the potential for similar overwintering plasticity combined with the plasticity observed in ants across a variety of traits (caste differentiation, queen-worker polymorphism, etc. positions supergene-carrying ants as interesting candidate systems for evaluating evidence of supergene-mediated genetic assimilation.

Outside of ants, chromosomal inversions, have given rise to supergenes that regulate complex trait polymorphisms in diverse taxa including flowering plants (e.g., Zhu et al., 2023), fish (e.g., Kirubakaran et al., 2016), birds (e.g., Lamichhaney et al., 2016), and butterflies (e.g., Joron et al., 2011; Nishikawa et al., 2015). The swallowtail butterfly (*Papilio polytes*) provides a compelling system for investigating potential supergene-mediated genetic assimilation. In this species, a ~130 kb autosomal inversion controls a wing-patterning polymorphism, resulting in both non-mimetic and mimetic forms. In an example of Batesian mimicry, the mimetic form resembles the toxic *Pachliopta aristolochiae* (Nishikawa et al., 2015). The inversion includes *dsx*, a known regulator of insect wing patterning, color, and structure (Nishikawa et al., 2015). Knockdown experiments demonstrated that the inverted form of *dsx* regulates the mimetic form (Nishikawa et al., 2015). The ancestral, non-mimetic form displays plasticity in response to temperature shock treatments, with the induced phenotype partially resembling the supergene-mediated patterning of the mimetic form (Shimajiri & Otaki, 2022). Other butterfly species (*e.g., Heliconius numata*) also display wing-patterning polymorphisms regulated by supergenes and exhibit plasticity with possible relevance to the evolutionary origin of potentially assimilated wing-patterning phenotypes.

Further characterizing teaching expertise: hypothesizing synergistic teacher knowledge and practice

Testing the quantitative student outcomes associated with teaching knowledge remains a high priority for future research in the field. However, since both Chapters 3 and 4 highlight these possibilities, I instead focus here on opportunities for future qualitative work to build on some of the findings of my teacher knowledge research. Specifically, I focus on a hypothesis that emerged from analyses of interview data and findings from prior work: via "synergistic teacher knowledge interaction", teaching strategies informed by one category of teacher knowledge may facilitate instructional goals related to other categories of teacher knowledge. Under this hypothesis, the outcome of such synergistic interactions is a more cohesive, expert-like body of teacher knowledge with knowledge components that support one another's enactment as opposed to presenting alternative goals that instructors must choose between.

Synergistic teacher knowledge and practice

Qualitative studies of undergraduate teacher knowledge have suggested that active learning expertise involves teaching strategies supported by one category of knowledge that enhance teaching goals related to another category (e.g., Andrews et al., 2019). Prior work revealed that instructors who frequently engage students in generative work drew on their pedagogical knowledge of how people learn and their PCK to design active learning instructional strategies that anticipated student difficulties with pedagogical supports to help them overcome those difficulties (Andrews et al., 2019). I also observed this in my data. For example, a participant deliberately left out information about the directionality of mRNA and tRNA at the ribosome during their lecture. This intentional omission was designed to give students the opportunity to grapple with a challenging aspect of the content during group work. This instructional strategy represented a possible synergistic interaction of pedagogical knowledge of how people learn and PCK in the design of an instructional strategy that accomplished pedagogical goals focused on engaging students in generative work in tandem with goals related to misconception-focused instruction.

However, in analyzing interview data from the studies presented in Chapters 3 and 4, other possible synergistic interactions between pedagogical knowledge categories and PCK emerged which I present below (Figure 5.1). Thus, this hypothesis expands on those prior to predict synergistic, bi-directional interactions between PCK, pedagogical knowledge of how people learn (HPL), and pedagogical knowledge of monitoring student thinking (MST). Chapters 3 and 4 focused on PCK and HPL respectively. Our interview protocol targeted MST knowledge, and Chapter 3 focused on this knowledge as a factor contributing to PCK development in case studies and the hypothetical model of PCK development. As a reminder, MST knowledge refers to an instructor's awareness of approaches to elicit detailed student thinking during class and the purposes of accessing student thinking and altering instruction accordingly (Andrews et al., 2019).

In Figure 5.1, I lay out the hypothesized synergistic, bi-directional interactions between each possible combination of teacher knowledge. In the following six paragraphs, I briefly provide one example of each type of synergistic interaction from my data. Pairing an underdeveloped hypothesis with an underdeveloped visualization has the potential to confuse the reader. To simplify, the same language and symbols have been used in each example to describe how a knowledge category (X) helps teachers implement an instructional strategy related to another category of knowledge (Y) “because they know how to” leverage their knowledge of the first category (X). Examples from my data aim to clarify the nature of these synergistic interactions.

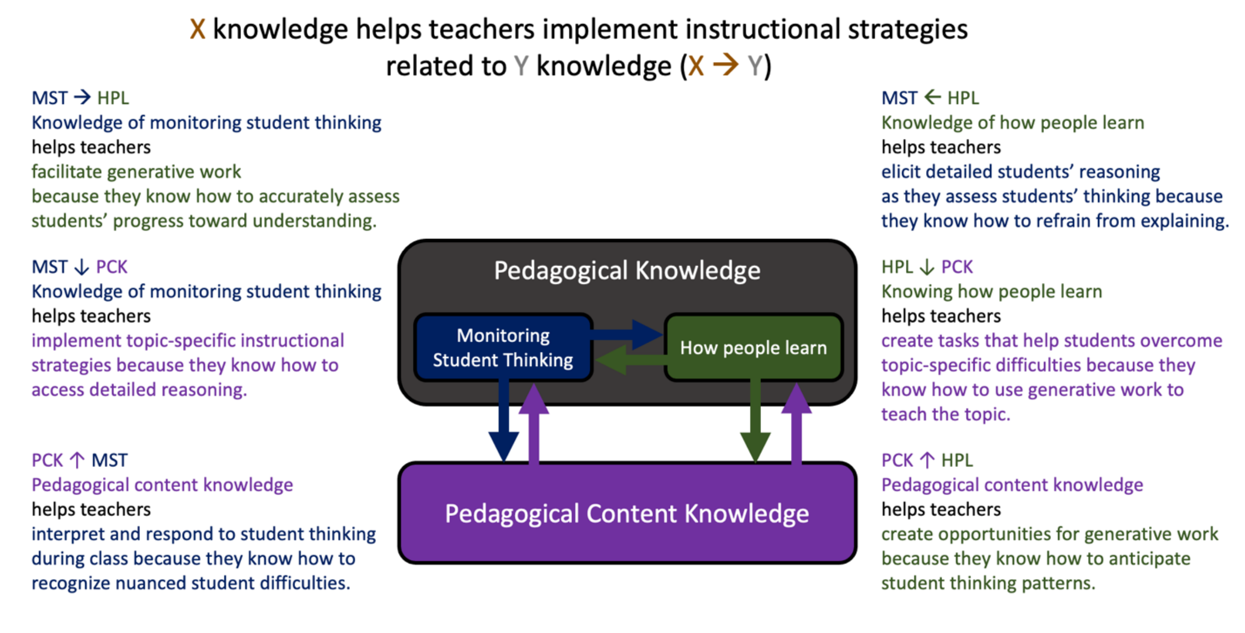


Figure 5.1. Hypothetical model suggesting how components of teacher knowledge support teaching practices related to other knowledge components. X and Y serve as placeholders to demonstrate how the knowledge and practice relationships can be here interpreted. X and Y can be replaced by any two of the three knowledge categories mentioned in this dissertation: pedagogical content knowledge (PCK), pedagogical knowledge of how people learn (HPL), and pedagogical knowledge of monitoring student thinking (MST). Examples explain how each knowledge component could support instructors in implementing teaching practices to support teaching goals not necessarily related to that knowledge component but pertinent to another.

MST helps teachers enact their HPL to facilitate generative work because they know how to accurately assess students’ progress toward understanding. In other words, knowing how to hear student reasoning, diagnose student difficulties, and respond at the scale of the entire classroom to patterns of student thinking that emerge. For example, Eric elicited student thinking during class which led him to recognize that students were struggling with a confusing visual aspect of one of his generative activities that was not relevant to the content or crucial for students learning. Eric demonstrated his MST as he recalled what he was thinking when he identified the issue during an interaction with students, "… just staying with that group of students. What's not clear, right? What are they having a confusing issue with? Oh, that's why it's not clear, because I made a mistake". Eric recalled that his immediate response was to think, "How do I fix this and still maximize the opportunity for them to learn?" Eric gave instructions to the class to correct the mistake, and he adjusted the visual the following semester to ensure any difficulties during students' generative work emerged only in relation to content and in support of their learning, a goal of HPL-related practices supported by MST knowledge.

HPL helps teachers enact their MST to elicit detailed students' reasoning as they assess students' thinking because they know how to refrain from explaining. Put another way, knowing that students, not the instructor, should take the lead in explaining during one-on-one interactions enhances an instructor's ability to assess the nuances of student thinking by hearing their reasoning. As we saw in her case from Chapter 4, Dana developed her knowledge of refraining from explaining, stating that "one of the challenges with this is trying not to talk, trying to wait it out. Because again, I can say it all I want, but I'm not the one who needs that practice verbalizing it. I already know what I'm talking about." This enabled Dana to refrain from explaining during her interactions with students which enhanced her access to evidence of student thinking for her to monitor, a goal of MST-related practices supported by HPL knowledge.

MST helps teachers enact their PCK to helps teachers implement topic-specific instructional strategies because they know how to access detailed reasoning from students. In other words, knowledge of how to elicit evidence of students' thinking helps instructors identify students' unproductive ideas which they can respond to with targeted instructional strategies. Henry wanted to identify the specific causes of and develop solutions for the difficulties students encountered with each topic. He also knew about the importance of eliciting and monitor student reasoning during class. For example, Henry told us about how he approaches his interactions with a student like a doctor treating a patient, “It's sort of like a doctor does at the doctor's office. Like you ask questions and based on how they answer, it's like a tree that you go down, right?... Okay, what did you do to get this? How did this happen? And just start asking questions to figure out, okay, what do they understand and what don't they understand and how can we get them back on track? Just kind of diagnostic questions.” This knowledge led Henry to take an approach to interacting with students that revealed insights into student thinking and yielded improvements to topic-specific instructional strategies, a goal of PCK-related practices supported by MST knowledge.

PCK helps teachers enact their MST interpret and respond to student thinking during class because they know how to recognize nuanced student difficulties. Put another way, PCK enables instructors to make sense of student reasoning as they shape their understanding of student thinking in real time (Gehrtz et al., 2022). For instance, Dana knew in advance that students struggled with thinking about sex determination in black and white, discrete terms, rather than as a product of more continuous phenomena such as relative gene expression levels. Knowing this in advance, Dana was able to ask questions as she elicited student thinking in her final semester to correctly diagnose a nuanced difficulty. Therefore, Dana's PCK allowed her to not only gather more evidence of student thinking during class but also to take a targeted approach in assessing the presence of specific difficulties among her students, a goal of MST-related practices supported by PCK.

HPL helps teachers enact their PCK to create tasks that help students overcome topic-specific difficulties because they know how to use generative work to teach the topic. In other words, knowing that students learn better via generative work equips instructors with a general principle to follow in teaching the difficult aspects of specific topics in their courses. For example, Irene knew that students struggled with distinguishing plasticity and selection as sources of phenotypic variation, but she needed an instructional approach to teach this concept. At first, she used a multiple-choice question, then she developed the HPL idea that “to explain is to learn, and this knowledge development corresponded with changes to the instructional strategy to make it an open-ended experimental design question. This strategy would be predicted by the ICAP framework to benefit student learning more than the multiple-choice question. Possibly by leveraging her developing ideas about how people learn, she developed a more effective topic-specific instructional strategy, a goal of PCK-related practices supported by HPL knowledge.

PCK helps teachers enact their HPL to create opportunities for generative work because they are aware of student thinking patterns. Put another way, awareness of student difficulties can increase the efficiency of an instructor’s enactment of HPL knowledge by giving an instructor a target at which to aim their knowledge of how people learn. In June’s case, she discovered that students struggled with a set of three challenging exam questions related to the lac operon. This PCK led her to implement these questions as an in-class activity the following semester. Instructors may be aware of generative work as an important process for learning but unsure about which topics in their lesson deserve the expenditure class time for generative work. Without sufficient PCK, they may incorporate opportunities for generative work to teach topics that students already understand or could efficiently learn from active or passive engagement. Therefore, PCK may enable instructors to engage students efficiently in generative work, a goal of HPL-related practices supported by PCK.

Future work could qualitatively assess this model in more detail using the data set presented in Chapters 2 and 3. If teaching expertise involves cohesion and synergy, then it is especially important to figure out not only how to help instructors develop in each of these categories of teacher knowledge, but to help them develop knowledge that synergizes across different categories to accomplish multiple pedagogical goals using individual instructional strategies. Similarly, the tensions we observed in case studies between teaching knowledge and practice could be due to an underdeveloped understanding of how multiple instructional goals can be achieved via single instructional strategies. For example, June's tension between her understanding of the importance of students explaining their reasoning and her tendency to default to providing explanations to students could be due to an underdeveloped understanding of how eliciting student reasoning accomplishes goals for both the teacher and student learning as it facilitates the instructor's PCK development, access actionable insights into student thinking during class, and student learning via generative work in one fell swoop. It could be important for instructors to recognize the multi-purpose nature of many instructional strategies, as this understanding can influence their decisions to use these strategies and encourage support-seeking behavior while learning to implement them.

Closing remarks

In this dissertation, I've shared my contributions toward understanding both the molecular mechanisms underlying plasticity in fire ant queens and the "plasticity" that human teachers demonstrate as they respond to their teaching environment to develop new knowledge and teaching practices. My evolutionary biology research has led to a rich characterization of the role of plasticity and supergene polymorphism in fire ants and identified the potential for an evolutionary link between the two. My teacher knowledge research has led to a rich characterization of the teacher knowledge instructors use to implement evidence-based teaching practices effectively and how they develop this knowledge.

Studying this "plasticity" in both fire ants and human teachers has left me with a unique perspective on how change can occur within an individual's lifetime. The change we observed in overwintered fire ants occurs as a response to an environmental stimulus. To put it simply, factors external to an organism can stimulate change within the organism. This phenomenon of phenotypic plasticity is widely studied in biology, occurs in the species *Homo sapiens*, and offers a useful metaphor that demonstrates a certain philosophical approach to thinking about change within a person. If one applies the concept of phenotypic plasticity philosophically, a person can understand changes in their thinking and behaviors as consequences of the environment influencing their internal state. This is a valuable perspective that acknowledges how surrounding material conditions affect the individual. Studying human change in the context of the development of their teaching knowledge and practices permits another perspective on change: internal factors can also stimulate behavioral change within an individual's lifetime that affect the external environment. Our participants developed new knowledge, often from self-reflection, that likely enabled them to improve their teaching for the benefit of students in their communities. Philosopher Jiddu Krishnamurti provides what I find to be a productive expansion on this idea by challenging the individual to look internally to affect external change in one's societal environment.

> *"To transform the world, we must begin with ourselves; and what is important in beginning with ourselves is the intention. The intention must be to understand ourselves and not to leave it to others to transform themselves… It is important to understand that this is our responsibility, yours and mine; because, however small may be the world we live in, if we can transform ourselves, bring about a radically different point of view in our daily existence, then perhaps we shall affect the world at large, the extended relationship with others."*
>
> – Jiddu Krishnamurti, *The First and Last Freedom* (1954)

References

Andrews, T. C., Auerbach, A. J. J., & Grant, E. F. (2019). Exploring the relationship between teacher knowledge and active-learning implementation in large college biology courses. *CBE—Life Sciences Education*, *18*(4), ar48.

Arsenault, S. V., King, J. T., Kay, S., Lacy, K. D., Ross, K. G., & Hunt, B. G. (2020). Simple inheritance, complex regulation: Supergene-mediated fire ant queen polymorphism. *Molecular Ecology*, *29*(19), 3622-3636.

Avril, A., Purcell, J., Brelsford, A., & Chapuisat, M. (2019). Asymmetric assortative mating and queen polyandry are linked to a supergene controlling ant social organization. *Molecular ecology*, *28*(6), 1428-1438.

Blacher, P., De Gasperin, O., & Chapuisat, M. (2021). Cooperation by ant queens during colony-founding perpetuates alternative forms of social organization. *Behavioral Ecology and Sociobiology*, *75*(12), 165.

Chapuisat, M. (2023). Supergenes as drivers of ant evolution. *Myrmecological News*, *33*.

DeHeer, C. J. (2002). A comparison of the colony-founding potential of queens from single- and multiple-queen colonies of the fire ant Solenopsis invicta. *Animal Behaviour*, *64*(4), 655-661.

Fontcuberta, A., De Gasperin, O., Avril, A., Dind, S., & Chapuisat, M. (2021). Disentangling the mechanisms linking dispersal and sociality in supergene-mediated ant social forms. *Proceedings of the Royal Society B*, *288*(1949), 20210118.

Gehrtz, J., Brantner, M., & Andrews, T. C. (2022). How are undergraduate STEM instructors leveraging student thinking?. *International Journal of STEM Education*, *9*(1), 18.

Joron, M., Frezal, L., Jones, R. T., Chamberlain, N. L., Lee, S. F., Haag, C. R., ... & Ffrench-Constant, R. H. (2011). Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature*, *477*(7363), 203-206.

Kay, T., Helleu, Q., & Keller, L. (2022). Iterative evolution of supergene-based social polymorphism in ants. *Philosophical Transactions of the Royal Society B*, *377*(1856), 20210196.

Keller, & Ross. (1993a). Phenotypic basis of reproductive success in a social insect: Genetic and social determinants. *Science, 260*(5111), 1107-1110. doi:10.1126/science.260.5111.1107

Keller, & Ross. (1993b). Phenotypic plasticity and "cultural transmission" of alternative social organizations in the fire ant *Solenopsis invicta*. *Behavioral Ecology and Sociobiology, 33*(2), 121-129. doi:10.1007/BF00171663f

Keller, L., & Ross, K. G. (1999). Major gene effects on phenotype and fitness: The relative roles of Pgm-3 and Gp-9 in introduced populations of the fire ant solenopsis invicta. *Journal of Evolutionary Biology*, *12*(4), 672–680.

Kirubakaran, T. G., Grove, H., Kent, M. P., Sandve, S. R., Baranski, M., Nome, T., ... & Andersen, Ø. (2016). Two adjacent inversions maintain genomic differentiation between migratory and stationary ecotypes of Atlantic cod. *Molecular ecology*, *25*(10), 2130-2143.

Lamichhaney, S., Fan, G., Widemo, F., Gunnarsson, U., Thalmann, D. S., Hoeppner, M. P., ... & Andersson, L. (2016). Structural genomic changes underlie alternative reproductive strategies in the ruff (Philomachus pugnax). *Nature genetics*, *48*(1), 84-88.

Levis, N. A., & Pfennig, D. W. (2016). Evaluating 'plasticity-first'evolution in nature: key criteria and empirical approaches. *Trends in ecology & evolution*, *31*(7), 563-574.

Nipitwattanaphon, M., Wang, J., Dijkstra, M. B., & Keller, L. (2013). A simple genetic basis for complex social behaviour mediates widespread gene expression differences. *Molecular Ecology*, *22*(14), 3797-3813.

Nishikawa, H., Iijima, T., Kajitani, R., Yamaguchi, J., Ando, T., Suzuki, Y., ... & Fujiwara, H. (2015). A genetic mechanism for female-limited Batesian mimicry in Papilio butterfly. *Nature genetics*, *47*(4), 405-409.

Rosset H, Chapuisat M. 2007 Alternative lifehistories in a socially polymorphic ant. Evol. Ecol. 21, 577–588.

Shimajiri, T., & Otaki, J. M. (2022). Phenotypic plasticity of the Mimetic Swallowtail butterfly Papilio polytes: color pattern modifications and their implications in mimicry evolution. *Insects*, *13*(7), 649.

Wang, J., Wurm, Y., Nipitwattanaphon, M., Riba-Grognuz, O., Huang, Y. C., Shoemaker, D., & Keller, L. (2013). A Y-like social chromosome causes alternative colony organization in fire ants. *Nature*, *493*(7434), 664-668.

Zhu, S., Zhang, Y. E., Copsy, L., Han, Q., Zheng, D., Coen, E., & Xue, Y. (2023). The snapdragon genomes reveal the evolutionary dynamics of the S-locus supergene. *Molecular Biology and Evolution*, *40*(4), msad080

APPENDIX

Appendix A. Pre-Lesson Interview Protocol

This is an interview with_____________ who consents to being audio-recorded.

The goal of this interview is to get a read on how you are thinking about the lesson we are going to film. There are no right or wrong answers to the questions I'm going to ask. I want to hear everything that comes to mind so I can understand the inner-workings of your teaching. Feel free to ask questions at any time.

CHANGES

We had the chance to discuss this lesson with you in [last data collection]. I'm particularly interested in any changes you made since you last taught this lesson, and changes that are not strictly about adjusting to COVID. What kinds of changes have you made?

[For each change THEY list].

Why did you make that change?

If not addressed, what prompted you to make that change?

If not addressed, what do you hope to achieve by making that change?

Then ask about changes you noted in the materials that they have not listed.

I noticed in your materials that you changed [any additions, omissions, changes to slides, handouts, cases, etc.].

Why did you make that change?

If not addressed, what prompted you to make that change?

If not addressed, what do you hope to achieve by making that change?

Remind me how many times you've taught this lesson.

Thanks for telling me about the changes to the lesson.

NOT a question: Establish learning objectives. Ideally, state the learning objectives that are shown in class materials. If not stated in class materials, ask what goals/learning objectives do you have for students in this class session?

1. Students can have productive and unproductive ideas about a topic. What ideas about [this topic] do students bring to this lesson? (PCK-ST)

. Why do you think students have those ideas?
   a. What makes you think that?
2. You've told me about [describe what they said about student thinking]. How does that influence your instruction? (PCK-IS)
3. What concepts and tasks do you expect students to have difficulties with in this lesson, and why? (PCK-ST)

   a. PROBE to cover all or most objectives/topics.
   b. PROBE to reach a point where they say “I don’t know.”
   c. What makes you anticipate these particular difficulties?
4. You talked about some of the difficulties you anticipate in this class. How does this particular lesson help students overcome the difficulties you anticipate? (PCK-IS)
5. Class seems to be structured as [describe specific instructional strategy]. Talk to me about the rationale behind this approach. (GPK)
6. Talk to me about how or why [specific instructional strategy] helps students learn. (GPK)
   a. Make sure to ask “why” or “how” until you understand their thinking about how students learn.
7. How do you communicate to students why you are asking them to [describe instructional strategy] during class? (GPK)
8. In what way is this [question/activity/discussions] particularly useful in helping students learn [specific topic]? (PCK)
9. During class, how do you know what your students are thinking? (MST)
   1. potential rephrasing if initial question fails: During class, what do you do to determine when students are struggling with [lesson topic]?
   2. Why is that [approach they use to monitor student thinking] useful?
   3. Why is it important to you to [approach they use to monitor student thinking]?
10. How do you hold students accountable for the work they do in class?
    a. Some things they may mention: Attendance, Work they do in class worth points, Exam alignment
11. Do you use specific strategies during class to encourage students to work?
12. Why is doing [ways they hold students accountable] important?
13. What else do you think I should know about this class period?
14. For COVID: We last interviewed you in [spring/fall]. When you think about the changes you made to this lesson, especially those we discussed earlier, what has influenced your teaching or your thinking about teaching since that time? Possible follow-up/clarification: Putting the technicalities associated with COVID aside…
15. For non-COVID: We last interviewed you in [spring/fall]. A lot can change in a [semester/year]. What has influenced your teaching or your thinking about teaching since that time?
    a. Probe about how experiences HAVE influenced teaching

   b. Probe about changes in thinking: And how do you imagine that might influence your teaching practices eventually?

16. And have you participated in any formal or informal teaching professional development since [last data collection]?
    a. When did you participate in this training?
    b. What did this training involve? (activities, intensity)
    c. How did this training inform your teaching?
17. Have teaching mentors or colleagues influenced your teaching in that time?
    a. In what ways?

Appendix B. Post-Lesson Interview Protocol
Clip selection

Guidelines for choosing videos: Keep two things in mind while choosing clips: the criteria below AND the list of required questions. For these criteria, meet criteria 1 first, then consider criteria 2 & 3 (which do not have to be met).

1) Choose at least 3 clips where the instructor gets access to student thinking by talking to students, seeing student work (including clicker questions), and hearing student questions (in order of importance).
   a) Prioritize clips that show students working on challenging, generative work (preferentially).
   b) If available, contrast that with clips where the work they are doing is more recall or simple application.
   c) Choose clips that show a diversity of how the instructor responds to student thinking, including asking prompting questions, explaining, providing info (in order of importance).
2) Choose moments where the instructor is thinking on their feet, including making decisions to change the instructional plan AND responding to student thinking (this will often overlap with criteria 1)
3) Choose moments when the instructor appears to be holding students accountable in a way that was NOT discussed in the pre-instruction interview.

Start by asking if they have any questions about the consent form AND get signatures.

Overview

Today we are talking about the class period where you … (remind them of what happened). We're not evaluating your teaching. We're aiming to understand the thinking that goes into your teaching. We're going to watch some clips from that class period, but before we do that I would like to hear your big picture thoughts about that class period.

ASK follow-up to understand their thinking about what they say.

Clips

Now we are going to watch five short clips from your class, one at a time. I want to know everything you were thinking as you were teaching. Just say anything and everything that comes to mind. I want to hear your running commentary of what you were thinking in the moment in class. I'll start the video and you can ask me to stop at any time. Or, as soon as you start talking, I will stop the video so you have a chance to tell me exactly what you're thinking.

Possible additional prompts:

- Try to get back in the headspace of this moment. What were you thinking?
- As a reminder, I want to hear everything you can remember thinking during this time in class.
- I would love to hear a running commentary of what you were thinking during this clip.
- What else can you recall thinking during that clip?

Use the following to build the clip-specific part of the protocol

Note: Before you play a clip, briefly describe the context in which the clip is taking place. For example, "This was the second clicker question you asked in the lesson." OR "This is when students are working on the back side of the worksheet." OR "Just before this clip, you had given students a minute to think with their neighbor about biodiversity."

These questions aim to elicit general pedagogical knowledge. Ask each one least once, unless there is no moment in the entire class that would allow you to ask it. ASK 9 minimum, including asking each question at least once (note that some questions have a necessary 'why' or 'how' follow up)

1. What are you trying to achieve as you interact with students in this moment?
2. In this moment, what were you thinking about your role in this interaction with this student? Why is that role important?
3. Tell me what you are thinking as you interact with this student.

4. In what way is this [general instructional strategy, such as think pair share, case study, worksheet activity] helpful for student learning?
5. What are you thinking about as you circulate the room?
6. What are you paying attention to in order to know if students are achieving the learning objectives?
7. What are you thinking as you view students' work/see responses to clicker questions in this clip?

These questions aim to elicit PCK. They need to be asked about specific learning objectives or topics OR about specific questions posed to students/examples given. ASK 7 minimum, including asking each question at least once.

1. Can you walk me through what you were thinking about this student's reasoning [about topic/question just shown in clip]? (KST)
2. What were you thinking at this moment about whether students are [getting topic X/achieving the learning objective]? How did you know that? (KST)
3. What were students easily understanding [about specific topic] and where were they struggling at this point in the lesson? How did you know that? (KST)
4. How did the ideas you heard/work you saw from this student [about specific topic] influence your next teaching move? (KIS)
5. In what way is this [specific question] particularly useful in helping students learn [specific topic]? (KIS)

Wrap-up reflections

Did anything unanticipated or unplanned occur in this class period?

What new understanding of student thinking do you have, if any, after teaching this class? If there is new understanding, what changes do you think that might lead you to make in the future?

What would you like to change the next time you teach this topic? Why?

What do you think worked well to teach [topic] that you will keep next time? Why do you think that worked well?

Is there anything else you'd like to add?