International Journal of Science Education

ISSN: 0950-0693 (Print) 1464-5289 (Online) Journal homepage: http://www.tandfonline.com/loi/tsed20

Doctoral conceptual thresholds in cellular and molecular biology

David F. Feldon, Christopher Rates & Chongning Sun

INTERNATIONAL JOURNAL OF SCIENCE EDUCATION, 2017
https://doi.org/10.1080/09500693.2017.1395493

Routledge
Taylor & Francis Group

Check for updates

Doctoral conceptual thresholds in cellular and molecular biology

David F. Feldon[a], Christopher Rates[b] and Chongning Sun[a]

[a]Department of Instructional Technology & Learning Sciences, Utah State University, Logan, UT, USA; [b]Center for Educational Innovation, University at Buffalo, Buffalo, NY, USA

ABSTRACT
In the biological sciences, very little is known about the mechanisms by which doctoral students acquire the skills they need to become independent scientists. In the postsecondary biology education literature, identification of specific skills and effective methods for helping students to acquire them are limited to undergraduate education. To establish a foundation from which to investigate the developmental trajectory of biologists' research skills, it is necessary to identify those skills which are integral to doctoral study and distinct from skills acquired earlier in students' educational pathways. In this context, the current study engages the framework of threshold concepts to identify candidate skills that are both obstacles and significant opportunities for developing proficiency in conducting research. Such threshold concepts are typically characterised as transformative, integrative, irreversible, and challenging. The results from interviews and focus groups with current and former doctoral students in cellular and molecular biology suggest two such threshold concepts relevant to their subfield: the first is an ability to effectively engage primary research literature from the biological sciences in a way that is critical without dismissing the value of its contributions. The second is the ability to conceptualise appropriate control conditions necessary to design and interpret the results of experiments in an efficient and effective manner for research in the biological sciences as a discipline. Implications for prioritising and sequencing graduate training experiences are discussed on the basis of the identified thresholds.

ARTICLE HISTORY
Received 31 August 2016
Accepted 18 October 2017

KEYWORDS
Biology education; conceptual change/ development; higher education; threshold concepts; doctoral education; resarch training

Doctoral training in the biological sciences is expected to prepare highly skilled scientists who can advance human understanding of the natural world. To meet these expectations, emerging scientists must both have mastery of established disciplinary knowledge and tools and be able to innovate within (and sometimes across) their respective areas of focus. Current research on the process of doctoral education predominantly examines the phenomenon through a lens of socialisation, in which students 'gain the knowledge, skills, and values necessary for successful entry into a professional career requiring an advanced level of specialized knowledge and skills' (Weidman, Twale, Stein, & Leahy, 2001, p. iii). These skills entail specific bench techniques, methodological and analytic

CONTACT David F. Feldon david.feldon@usu.edu

skills, and application of theoretical frameworks. However, the socialisation research examines primarily students' 'process of internalising the expectations, standards, and norms' of their disciplines (Austin & McDaniels, 2006, p. 400) and focuses on the development of a scholarly identity and ability to engage in appropriate discourse within an area of specialty (Lovitts, 2008). Most studies that focus on research skill development specifically examine feelings of preparedness for conducting independent research (e.g. Austin, 2002; Delamont & Atkinson, 2001; Golde & Dore, 2001; Lovitts, 2001; Pole, 2000). Other studies focus on academic productivity measures as proxies for evidence of skill (e.g. Paglis, Green, & Bauer, 2006). However, given the high involvement of colleagues and mentors in the development of such work products, it is not possible to derive specific information about the relative strengths and weaknesses of an individual student's skills or scholarly growth other than through personal opinion (Feldon, Maher, & Timmerman, 2010; Isaac, Quinlan, & Walker, 1992; Lipschutz, 1993).

Consequently, very little is known about the mechanisms by which discrete competencies are acquired and how specific practices on the part of doctoral advisors or Ph.D. programmes could most effectively leverage them (Gilbert, Balatti, Turner, & Whitehouse, 2004). Even in the postsecondary biology education literature, identification of specific skills and effective methods for helping students to acquire them are limited to undergraduate education (e.g. Boyer, 2003; Brewer & Smith, 2011; Caldwell, Rohlman, & Benore-Parsons, 2004; Labov, Reid, & Yamamoto, 2010). Several studies have examined strategies used in experimental design generally (e.g. Feldon, 2010; Schraagen, 1993; Schunn & Anderson, 1999), and others have examined the impacts of specific graduate education practices on research skill development across science, technology, engineering, and mathematics (STEM) disciplines (e.g. Feldon, Shukla, & Maher, 2016 ; Feldon et al., 2011). However, these have not been able to inform the practice of graduate education in a targeted way, because they have not focused on these phenomena within a single discipline. Disciplinary context is an important element of socialisation, because it frames an intellectual identity, parameters for relevant and appropriate research questions, and accepted research methods (Chubin, 1983; Kuhn, 1962; Price, 1965), summarised by Knorr-Cetina (1997, p. 260) as 'epistemic cultures.'

In order to establish a foundation from which to investigate the developmental trajectory of doctoral students' research skills specifically within cellular and molecular biology, it is necessary to identify those skills which are integral to doctoral study and distinct from skills acquired earlier in students' educational pathways. Without such a foundation, future efforts to inform educational practice within the discipline will be hampered by potentially overbroad generalisations from prior research or insufficient differentiation from undergraduate skill development.

Threshold concepts

Few studies have focused specifically on the development of students' knowledge and skills as they evolve over time within the context of graduate education. However, Kiley and Wisker (2009) suggested that graduate students mature into independent scholars through the acquisition of threshold concepts, which 'once grasped, lead to a qualitatively different view of the subject matter and/or learning experience and of oneself as a learner' (p. 294). On the basis of interviews with university faculty, they characterised the cognitive development of graduate students as they transition from disciplinary novices to independent scientists as a progression of threshold concepts. These concepts are described as 'so critical to an understanding of the discipline that advanced disciplinary learning is not possible … [until] the threshold of understanding for that concept [is crossed]' (Kiley, 2009, p. 297). Attainment of these skills was considered to be neither gradual nor linear and required mastery of certain knowledge or skills prior to the attainment of others. Specifically, thresholds concepts are understood to be transformative, integrative, irreversible, and challenging. As such, threshold concepts are both obstacles and significant opportunities for intellectual gain.

Threshold concepts are not considered equivalent to core concepts of a field of study (e.g. biology). Core concepts serve as building blocks that increases a person's understanding in a subject, but they do not inherently lead to a different view of the subject or greater ability to conceptualise and solve problems in the discipline (Meyer & Land, 2003). For example, learning about individual species within a genus may expand a biologist's understanding of the field, but the additional information does not change how the biologist conceptualises the broader framework of the discipline or focus of research. In contrast, a threshold concept, once understood, qualitatively alters one's view of a subject or discipline, one's learning experience, and/or one's identity in relation to a discipline (Meyer, 2008).

Threshold concepts in the context of doctoral education

Kiley (2009) identified at least three major threshold concepts in the development of doctoral students' scientific reasoning. These threshold concepts, elicited through extended interviews with faculty who mentor doctoral students, focused on the nature of empirical research and the criteria that distinguished high quality research from efforts that lack essential disciplinary indicators of rigour. First, students must construct an 'argument or thesis supported by defensible evidence' (p. 298) that accommodated or reconciled conflicting views and data. Second, they must have generated a theoretical model that allowed findings to be applicable in other situations. Third, students should have been able to articulate an awareness of the conceptual framework or intellectual context from which the work arose.

The first possible threshold concept, *Argument*, required that arguments or theses must have been supported by defensible evidence (Kiley & Wisker, 2009). A second proposed concept, *Theory*, required students to generate a theoretical model that allowed findings to be applicable in other situations. This proposed concept required students to demonstrate their ability to integrate information into other areas of their discipline and therefore transform their understanding in those areas as well. The third potential threshold concept, *Framework*, required students to articulate an awareness of the conceptual framework or intellectual context from which their work arose (Kiley, 2009). This entailed conceptually locating and bounding one's research in the relevant primary literature.

From the perspective of performance-based assessment, Timmerman, Feldon, Maher, Strickland, and Gilmore (2013) provided convergent support for these thresholds using a sample of graduate students from across STEM disciplines. By examining criterion-referenced scores from graduate students' written research proposals, Timmerman and colleagues demonstrated that students' ability to contextualise their proposed studies in the context of current work in a scientific field and appropriately use primary literature reliably preceded other aspects of research skill development, consistent with Kiley's (2009) identification of *Framework* as a threshold concept. However, other observed patterns of skill development indicated that *Argument* (Kiley & Wisker, 2009) may not differentiate sufficiently between more nuanced scientific skills that developed independently at differing points in students' learning trajectories. Less experienced students demonstrated substantial growth in developing testable hypotheses, whereas the ability to effectively ground conclusions in data improved substantially later in students' development.

Research questions

While helpful as orienting concepts, currently defined threshold concepts are limited in their ability to inform practice within specific disciplines. To date, scholars have not identified the extent to which they operate universally or differ based on a student's field of specialisation. Given the canonical framing of threshold concepts as domain-specific (Meyer, 2008), the current study is positioned to better inform graduate education practice in cellular and molecular biology by identifying concrete areas of transformational difficulty for future scientists in this field. If able to determine the manner and extent to which such threshold concepts apply for these students, their doctoral programmes may be able to increase the efficiency and effectiveness of training by sequencing student training experiences to target the attainment of threshold concepts appropriate to their development. Thus, our research questions are:

(1) What core competencies do Ph.D. students in cellular and molecular biology experience as threshold concepts during their graduate training?
(2) What features do these threshold concepts have that may differ from broader frameworks in relation to the specific disciplinary context?
(3) What sequential transitions are experienced or described during the acquisition of identified threshold concepts?

Methodology

There is very little that is known about which skills may effectively function as threshold concepts within graduate cellular and molecular biology education. Therefore, this is an exploratory study meant to discover how students defined skill development at different stages within their Ph.D. programme and how they perceived their identities to develop in relation to this. We employed an interpretivist paradigm (Erickson, 1986) to engage the meaning and perspectives that current students and recent graduates expressed regarding their doctoral learning experiences in cellular and molecular biology: what they learned, how they learned it, how the experience affected their identities as researchers, and how they interpreted the experiences of their peers in their doctoral programmes.

Context

Cellular and molecular biology is 'the study of the structure, function, and behavior of cells' (Alberts et al., 2014, p. 1) in which the unit of analysis may vary from whole cells to individual molecules that enable a given cellular function. Because these phenomena occur at a scale too small to be observed by the human eye, research in this field occurs in a laboratory environment and utilises a number of highly specialised tools and experimental techniques to collect and analyse data about these functions. Such approaches include microscopy (imaging), centrifugation, electrophoresis, and chromatography (separation and isolation of cellular components or products), spectrophotometry (measuring light absorption to identify the properties of components or products), and DNA sequencing (determining the order of nucleotides within a DNA molecule). To appropriately interpret the data obtained, values from a target sample are often experimentally compared against values from control samples that differ in planned ways to isolate sources of variance (Mahmood & Yang, 2012).

In the U.S., doctoral education typically entails extensive supervised research experience in a professor's laboratory and serves as the primary mechanism by which individuals are trained to conduct this research. Such environments typically impose high demands on students' time (~45 h per week; Ferreira, 2003) and impose substantial pressures to quickly master both concepts and techniques for the purpose of contributing to the publications and funding proposals produced (Alberts, Kirschner, Tilghman, & Varmus, 2014; Kleinman, 1998). Within this context, responsibilities for supervision and mentoring of doctoral students are typically distributed in a 'cascading model' (Golde, Conklin Bueschel, Jones, & Walker, 2009, p. 57), wherein senior Ph.D. students most often receive much of their training from postdoctoral researchers within the lab rather than directly from faculty. In turn, senior students provide assistance to less experienced students within the lab.

Data collection methods

Participants were recruited through two mechanisms. First, recruitment materials were disseminated through email listserves maintained by the Society for Integrative and Comparative Biology, the Society for the Study of Evolution, the RNA society, the American Society for Cell Biology, and the American Society for Microbiology. Second, emails were sent directly to cellular and molecular biology professors at universities in the U.S. designated as 'research intensive' (R1) by the Carnegie Foundation for the Advancement of Teaching. We used a purposive sampling strategy to recruit participants with diverse experience levels from 17 institutions of varying size and geographic location.[1] These institutions included both public and private universities located in the Western, Midwestern, Southern, and Eastern U.S. (see Table 1), and enrolments in STEM discipline Ph.D. programmes during the year of data collection ranged between 478 and 6495 (National Center for Science and Engineering Statistics, 2017). In order to benefit from the broadest

Table 1. Participant information by geographic region and participant position.

U.S. geographic regions	Student	Postdoctoral researcher	Faculty
East	10	3	0
Midwest	4	3	0
South	5	9	4
West	5	0	1

personal perspectives, current, prospective, and retrospective, participants were selected who were currently enrolled in doctoral programmes, as well as those who had completed their programmes within the previous five years (e.g. postdoctoral research associates, assistant professors).

Thirty-two interviews were conducted over the phone, and 12 interviews were conducted in person. Participants received $20 as a research incentive. Interviews for all participants followed a script meant to elicit opinions about individual experiences during Ph.D. programmes. The script included background questions about their prior education and general areas of research interest, as well as focal questions such as 'What are the most important research skills or concepts that you learned during your Ph.D. programme?', 'Which skills or concepts were the most difficult for you to learn or understand?', and 'Where were the points where you felt stuck in your development as a researcher, and what understandings helped you to move forward from those points?' Interviews took approximately 1 h each and were deemed complete when the interviewer had exhausted responses from interviewees. All interviews were transcribed, and interviewer's maintained field notes to guide data analysis.

In addition, two focus groups were conducted with students and postdoctoral researchers from laboratories at the institutions where in-person interviews were conducted (one in the East, one in the West). Lunch was provided to participants as an incentive. These conversations attempted to elicit divergent perspectives (i.e. disconfirming evidence or added variance in participants' perspectives) on training experiences and potential threshold concepts by first explaining the idea of a threshold concept and sharing prospective examples drawn from previous interviews. Focus group participants were asked to comment on them, articulate the extent to which their own experiences were similar or dissimilar, and add any other possible threshold concepts from their own experiences. Following this, they were asked to discuss which, if any, of the concepts or skills discussed they might consider to be transformative, what sequences and contingencies might exist among the discussed skills, and by what criteria those skills might be recognised or assessed. In all cases, participants were encouraged to discuss their perspectives and experiences (both converging and diverging) with each other freely as part of the focus group.

Data analysis

As interviews progressed and possible threshold concepts were generated, we began to test these assertions at the end of interviews to find out if other students had had similar experiences. Furthermore, as themes began to develop, we noted if other participants reported dissimilar or divergent experiences.

Data analysis consisted of a mix of induction and deduction (Erickson, 1986). To identify possible threshold concepts related to research, we looked for signs that a particular idea or skill was integrative, transformative, or troublesome (per Meyer & Land, 2003). A particularly helpful basis for identifying threshold concepts by Davies and Mangan (2005, p. 7) was to look for concepts that were:

> Integrative in that a student who fully understands this concept will recognise and use this same tool in different areas of the discipline, and it may be transformative in that

understanding gained from use in a particular discipline area may lead to deeper understanding of a topic in a very different area.

We also looked for areas where students were either troubled in their advancement or felt they were in a liminal state (i.e. 'a period of uncertainty, confusion, or doubt' [Keefer, 2015, p. 18] or 'stuckness' [Kiley, 2015, p. 53]) and unable to move forward in their research or scholarly development. While any of these characteristics do not inherently indicate a threshold concept, they were starting points from which to focus our coding.

First, transcripts were read holistically and independently by each of the authors to get a sense of the variety of experiences students had and to create general categories of areas where threshold concepts may have been indicated. Each interview and accompanying field notes were then broadly coded by the second and third authors. Through iterative discussion amongst all authors, codes were compared and synthesised until consensus was reached in order to develop finer grained competencies that might function as threshold concepts. Assertions were then generated and evidence from interviews and the literature were used to support these in the results below.

Validity criteria. In order to help protect against threats to validity we employed several methods commonly used in qualitative research. Interviews were recorded and transcribed in order to make sure that as observers we correctly captured what interviewees said and to have a basis for what was meant in statements due to inflection and tone. To make sure that potential threshold concepts were not cherry-picked, only those that were coded in multiple interviews and observations were used to generate assumptions. Triangulation was further used to understand the variety of experiences around single competencies. Disconfirming evidence was also considered both during interviews (students were asked for alternative explanations) as well as during coding (students who had different experiences with the same process were integrated into the analysis).

Findings

In this study, two robust concepts emerged as candidate threshold concepts from the analysis of interviews of graduate students, postdoctoral research associates, and early-career faculty with remarkably little variation. First, participants reported that graduate students' earliest and reportedly most important struggles related to appropriately engaging the literature as a dynamic and fallible set of claims that contributed to larger discourse within the field. Second, participants discussed extensively the challenges and necessity of mastering the appropriate selection and use of positive and negative controls in designing experiments. This theme reflects a subset of experimental design skills specific to biology (e.g. Gross & Mantel, 1967). Of the 24 students, 15 postdoctoral research associates, and 5 early-career faculty who participated in interviews, only two participants reported that they had not experienced something that they could identify as a prospective threshold concept. Both were students who could not identify any conceptual shift or specific turning point (e.g. 'I don't think I've had any epiphany'; Tara, third year student). Similarly, the vast majority of focus group participants likewise identified either having crossed or struggling with

Table 2. Code frequencies by students, postdoctoral researchers, and faculty.

Themes	Student	Postdoctoral researcher	Faculty
Literature	19	4	4
Primary literature	13	3	4
Literature critiquing	13	2	2
Balanced views of literature	7	1	2
Controls	14	1	5
Design with proper controls	14	1	4
Challenges in selecting controls	4	0	3
Conceptual transformation	4	1	1

either or both of these threshold concepts. The number of interview participants who articulated experiences aligned with each theme and its major characteristics is presented in Table 2.

Primary literature

Primary literature refers to engagement with the foundational and current literature that summarises what is known about a discipline. This is where students learn the context of their field, where they can contribute meaningfully to the discipline, and why that contribution will be important. Primary literature provides a portal for beginning students to join the larger conversations within a scholarly community (Maher, Timmerman, Feldon, & Strickland, 2013; Urquhart, Maher, Feldon, & Gilmore, 2016). As Kimberly and Ron, both assistant professors, explained:

> Every student is different and what I do is typically for those kinds of students [who are new to the field] I would hand them … the ten most important papers for our field. By reading these you'll be able to have a basic conversation with anybody in the lab. When students leave, I want them to be able to know how to read the literature and find a hole. How to start to dig in, to fill that hole with knowledge in terms of science, and the tools and technology that are available. And then how to interpret that finding that then opens new doors that opens new holes and new pathways. (Kimberly, assistant professor)

> I would say the number one thing in terms of a theoretical base is reading the key papers. I mean, I'm talking about really reading 'em and criticizing them and kinda tearing 'em down, appreciating what's good about them and what's bad about them, and maybe why there were particular limitations. (Ron, assistant professor)

As detailed in the sections below, development in the ability to effectively utilise primary literature was often described as moving from a passive acceptance of the literature to excessive critique, and finally to a more dynamic ability to analyse strengths, weaknesses, and relevance to specific needs. Students came into their programmes with some understanding of their fields, but not enough knowledge of either procedures or content to derive their own implications for published work. They had to be told by advisors or senior peers what the results of experiments meant and why these were important. As students read more articles, learned more about their field, and saw how others in their field approached the literature, they were able to read strategically and independently determine what a given study could say about its phenomenon of interest within the broader field of study.

Initial exposure to current empirical literature was generally considered challenging.

> I remember it being very difficult initially. It used to take a long time to read a paper at enough depth that you could really understand all the aspects of what they were doing; understand the motivation behind all the experiments they did. (Greg, postdoctoral researcher)

> That was a big thing to hit me was the literature is incredibly diverse. … It's all very, very, very different, and there's a lot of sort of thought and experience and training that goes into just being able to take a paper and understand it. (Josh, second year Ph.D. student)

> I don't think you learn how to read until you've done it several times. And I think you need someone to tell you how in order to do it. There's a lot in science that is up for interpretation and it makes it hard to get the gist of a paper quickly … I think I had to figure out the question that they were addressing and I wasn't picking up on that. I wasn't focusing in on key areas that they were talking about. I think when I was reading the results I wasn't actually looking at the figures, I was looking at what they were saying about the results. What they say in the results may not be what they actually show. It's their interpretation of their results. I think by looking at the figures and saying here's the experiment they did, here's what the experiment shows, and here's what they say from that … I was reading the entire thing, and I wasn't grasping what they were trying to show. (Jeff, postdoctoral researcher)

Another postdoctoral researcher called this 'taking the literature for granted.' This was commonly described both as reading an article from cover to cover, as well as passively accepting conclusions reached by the authors. Carol reflected on her own approach as she first started reading literature.

> I came in with this idea that I'd read this and it's the word of God, and I'd get these academic crushes on people that wrote about topics that I was very interested in. I just thought their work was brilliant. To some extent most of that ended up being true, but I had been thinking about it for a long time and had done a lot of reading. And I read the discussions, and they tell you what's good and bad about their own study. And you follow up on that … . I didn't actually start criticizing in a negative way until we had those group discussions.

Reading passively and taking everything for its face value, as Carol did, was a common theme that resounded through our interviews with faculty and student participants. However, this idealism that everything they read was 'all being factual and carved in stone, that's the way it is' (Shirley, Ph.D. student) diminished as they gained more knowledge and experience.

Learning to read: developing a critical eye

Participants described learning to critique primary literature as requiring a long gestation period, during which they acquired knowledge of their field piecemeal well before they could integrate that information into a cogent argument with or about a publication. One student explained the length of the progression towards critical reading of the literature in this way:

> I would say it took some time to get to the point where I could actually read something and then critique it … . What helped me were courses … where we were getting general articles, critiquing them … . I think those courses, along with me spending my own individual time, helped me get to that point.

Shirley (doctoral student) explained that claims presented in the literature are 'sometimes … opinions, not necessarily hard, cold facts.' Ciara (fourth year doctoral student) added that 'part of being a scientist is questioning if [the authors] have made the right conclusions from interpreting their data, because sometimes their data could be interpreted in multiple ways.'

For example, Debora, a newly admitted doctoral student and former lab technician, asserted:

> I think once I realized that critiquing a paper is important, I started focusing more on trying to evaluate the data myself rather than just reading what the authors had to say about the data. Before, I would really try to understand how the authors extrapolated their conclusions from the data what was presented as opposed to me trying to evaluate the data on my own. I think now what I try to do more is I will look at exactly how the experiments were designed, and then I look at what the data tells me, and then I ask whether or not the conclusions that I draw are similar to the conclusions that the authors draw.

This more critical perspective also translated into a stronger ability for students to detect flaws in their own lab work. Kimberly (assistant professor) observed that when her students were able to read critically,

> It changes the way they're looking at their own work and it alters the path that they would have continued on. Instead of continuing to do an experiment a certain way with a certain control, they change the way their science is happening. Different controls are used. Or more controls are used. Or they start using controls. It's a change in the behavior, and the way they're thinking and performing science.

Similarly, Jason and Kelly (Ph.D. students) elaborated on how it benefitted their research:

> I think that [being critical of your own work] just helps going forward so if you can identify what shortcomings in your own work, that's something that needs to be addressed. You got a better idea of everything that needs to be done to kinda make your study complete. (Jason, senior doctoral student)

> Once you can be critical of the literature … you can then start to look at your own work in the same way. So that you can go, okay, somebody is going to look at this critically down the road. You start to analyze your data just as extensively as you do the stuff in the literature, so it really helps you kind of critically look at your data and how it's being done. (Kelly, senior doctoral student)

Balanced views of literature

Although more advanced students realised they were supposed to be critical of papers, they did not have a deep understanding of what that meant from an expert perspective. Students sometimes engaged mimicry of critical approaches to reading, which failed to provide the most productive perspectives on the literature they read.

Once novice students realised that the papers were fallible and their 'academic crushes' on those who published the papers were dispelled 'after a year or two' (Kelly, Ph.D. student), they often become hypercritical of all papers. 'And what happens, they go from thinking everything is perfect law, and then all of a sudden they become super critical. And there's about 6 months to a year where nothing is good enough' (Kimberly, assistant professor). Carol confirmed this from a student's perspective:

> Well, learning everything was wrong probably took 3 or 4 sessions of group discussion … I heard everyone criticizing these papers that I thought were really cool and then I realized I didn't understand what was going on and needed to read more carefully and really

> understand it, and that took a long time. It didn't take me very long to find what was wrong, but … then I just questioned everything … We actually had one girl in our lab … [who] wouldn't actually read the entire paper; she'd just find one thing that was wrong and just say it over and over again. Not getting a full understanding for what are the good parts of a paper, she just found what's wrong with it.

Carol also described her 'maturity' in balancing the interpretation of strengths and weaknesses within published articles:

> You can always find something wrong. It's much harder and much more of a challenge, but it also shows you that you're doing better if you can also find the things that are good about a paper … It took probably another 3 years before I really had a handle on being able to find the good and bad in something … it's what we all do when we criticize a paper. We disagree about what's good and what's bad about it, but my ability to do that matured … I think the ones that were better knew a little bit better how to criticize and see the good in things … It was kind of a three step thing. I had that idealism, then oh crap everything is wrong, then I had to gain more experience to see those papers again for what was good in them.

This pattern was described slightly differently by two other assistant professors. Ron and Matt viewed student development in reading primary literature and also engaging students' abilities to self-critique. As summarised by Ron:

> It is an interesting but stereotyped progression that they go through. Let's be hypercritical of everything; that's normal. Then let's start being hypercritical of ourselves. Then let's learn to take criticism from other people. Then let's then start being a little bit more forgiving with other papers; start appreciating the bad things and the good things.

Designing experiments with proper controls

As students talked about the ability to productively engage primary literature, they also began to reference their development as scholars with an ability to make their own contributions to it. It was in this context that skills highly specific to the biological sciences emerged – specifically the ability to appropriately select controls during the process of designing experiments and using them efficiently during the analysis of obtained data.

In the interviews, many participants emphasised the crucial nature of controls to an experiment. As Danh, a recent doctoral graduate, summarised, 'the good scientist is the one who designs better controls.' Poor controls often render an experiment unusable. Jason, a fourth year doctoral student, recalled of his early experience with unsuccessful experiments,

> Getting the data that was hard to interpret was I think probably mostly due to having poor controls built in. When you're looking at your results [and] the controls were poorly designed, then the data becomes harder to interpret … . I see a lot of situations [when] we have to trash [the experiment] because the control doesn't work.

Picking appropriate controls is usually troublesome for novice graduate students. Christina, a recent graduate, shared her experience with designing experiments when she started her programme:

> I think designing the experiments was hard for me to learn in that I did a lot of experiments and then realized, 'oh, I should have done this step differently, or this control wasn't the correct control to include', that sort of thing. A lot of times it was, 'oh, I forgot to do this

control' … I guess I wasn't very clear [about controls], but I just realized basically with designing experiments you only want to change one thing in order to look at how that – what the effect of that change is.

Heidi, a first year Ph.D. student, agreed and explained her difficulties with controls:

I would say that controls are probably one of the hardest things to do well and consistently … and I definitely struggle with this. It's like you're so eager to get to the results that [you think] the controls are things that you already know, and those don't matter. But it's impossible to really understand something unless you have it controlled well. I mean, controls are the hardest things to do consistently and well, and I definitely have a hard time with it.

Challenges in selecting controls

Although most participants reported having laboratory experiences prior to beginning their programmes, their previous research experience did not effectively prepare them to engage effectively in experimental design. As undergraduate researchers, most of them worked under the supervision of a senior lab member and had least, if any, control over the experimental procedures. Furthermore, in most cases, participants had worked on experimental protocols for which the controls were predetermined. Thus, designing controls was unique to their graduate school research experiences. Byron, a junior faculty member, recognised that beginning students, depending on their background, show variant patterns of deficiency in using controls. When asked if students were aware that they should be using controls when they first started or if the difficulty was better described as not knowing which controls to use, he explained:

It could be any combination. They're just not sure. It depends on their background too really. If they have no experience with actually doing bench work, a lot of times, yeah, they're gonna have problems knowing any controls or that they really need controls. They probably know in a sense [that] they do, but once they actually set up the experiment, 'let's do it', 'Oh, I completely forgot about controls or what I needed to do.' Someone that has experience probably knows that there needs to be controls; although they may not know exactly what the control is and whatnot.

Participants expressed three aspects of selecting and designing controls that challenged students: control type, quantity, and quality. There are two types of controls in biological experiments (Ruxton & Colegrave, 2010). Positive controls confirm an anticipated effect of an experimental mechanism (e.g. verifying that an antibiotic is effective at killing bacteria as expected). Negative controls confirm that there is no effect where there should not be (e.g. two different strains of bacteria can coexist in a petri dish without negatively impacting one another). Thomas, a faculty member, observed that 'the vast majority of experiments of junior graduate students do not include positive controls. If they have control, they have a negative control. The negative control is no drug, right, where nothing should happen.' Byron concurred:

I think negative controls, for whatever reason, they're the ones that people mostly think about … . Yeah, there's obviously definitely positive controls that they have to do a lot of times as well. For some reason, it seems to me that most people tend to pick up on the negative controls.

Carol reported that she had often struggled but was able to do research more easily once she understood the use of both positive and negative controls:

> One thing is the importance of controls, doing an experiment, and then after the experiment, and then realizing you've no idea that you've been contaminating something because you didn't have the right negative controls, also having the right positive controls for something like a PCR (polymerase chain reaction, used to replicate large quantities of a specific segment of DNA), something where you're not sure if you screwed up or if it's just not there because you can't see it.

When making decisions about the quantity of controls, beginning students are more likely to add more controls than necessary into an experiment, making an already complicated study more complex. Thomas explained:

> A typical experiment for an incoming student often – there'll be ten treatments, minus, plus, this, that, and the other thing … So, at the end of the day, they have 40 different things that they're looking at. People end up designing these experiments that just end up being hopelessly complicated. Then the experiment's a big failure.

Kimberly, another faculty member, agreed: 'This is where controls are always the students' worst nightmare, because [as] faculty, we can come up with more controls [when critiquing a study design].' However, a hallmark of research expertise is also the ability to select the most productive controls. For example,

> some of the controls – comparing mutant to wildtype, it's a control, but it's probably not the most informative control. You're basically just saying that apples and oranges are different. So I think being able to design an experiment's controls, the correct controls that are the most informative … Quality is more important than quantity.

As Kimberly observed from her work with graduate students, '[u]sually they go through that phase of no controls, to lots of controls … and then learn to pick the ones that are most important for the experiments that they're doing.' During the no-control phase, as Byron mentioned previously, students were somewhat ignorant of controls – they either included no controls or had the sense of using controls but did not know what controls to use. After a couple of experimental failures, students learned the importance of using controls and started to be cautious of the control conditions. However, this newly gained insight tipped the students over to the other side of the scales – they demonstrated an inclination to design more controls than necessary. Christina, a fourth year doctoral student, shared with us an anecdote where she and lab mates were thrilled about an amazing result of her experiment, but it turned out she had changed the control conditions without knowing it beforehand. Thus, 'that result wasn't as exciting as we had thought for a little while.' To conclude her story, she confessed:

> With that example I … gave you it was really exciting and then it was really disappointing, and I felt really stupid. I think it made me more careful in my planning – especially with the more tricky or the more detailed experiments. [N]ow I guess I try to err on the side of having more controls than I need rather than less.

The tendency to use too many controls indicates basic progress away from not recognising the need for controls at all. However, it also indicates a possible liminal state in the course of learning to use controls, where they are stuck with the illusion that more controls means better design and, more importantly, being intelligent enough and knowledgeable enough for their programme. Kimberly made an illustrative note of this phenomenon:

> So the students will go through no controls, then uber-controls, where they control fourteen different ways, and wasted four months in the lab. … They get to a point where they realize ‘I could have just done those two experiments,’ because they showed between the two of them all fourteen things the other experiments showed … and then they realize ‘I don’t want to do that ever again.’ So they narrow down and realize which are the most elegant. When they design new experiments instead of using ten different neuronal markers to count cells, they’ll pick the two that are pan-neuronal.

However, Ron noted that this progression is not universal. ‘I’ve had some students who were just really good at controlling experiments, and I’ve got some students … that just never really pick it up.’

Conceptual transformation

Once students passed the threshold controls inflicted, their research lives became easier and more productive. Mastery of controls made research easier, ‘because you’re not having to go back and repeat experiments because you haven’t put the proper controls in place to begin with,’ as Kelly, a fourth year student, stated. In addition to precluding the nuisance of repeating every experiment, the mastery of controls also transformed students’ understanding of experiments. Carlo stated:

> I have to say that the thing that after I learned made my life a lot easier was determining the appropriate controls for each experiment because in the end, everything we do is a comparison … . The hardest part for me was always when I was designing my experiments was to think about what were the appropriate controls and once I got them down and I said ‘okay so for this animal model, I need to control for this, for this and for that,’ then my life got a lot easier … . Once I got down how to appropriately control my experiments, then the whole process of gathering data for my Ph.D. became a lot easier because I could do better comparisons and then I could get better data.

Similarly, Scott observed:

> In order to interpret the results of the experiment, you have to have a rock solid, airtight understanding of [the] control situation. If you really understood the control … then when you come to the experiment … if anything goes awry, then you can pick it out immediately.

Carlo further noted the characteristic difference between knowledge of techniques and the crossing of a threshold concept regarding the effective use of controls:

> The hardest part for me when designing my experiments was figuring out appropriate controls … . Once I figured out the controls, the whole experimental part got easier. The technical part is just practice, that didn’t change – if you know how to do a western blot (a protein detection technique) for experiment 1 then you know how to do it for experiment 2 – but the experimental designs and models will always be different.

Conclusions

In this study, we found two potential threshold concepts: (1) critical but balanced reading of primary literature and (2) the design of experiments with disciplinarily appropriate controls. Although there were a variety of paths towards the first threshold concept of balanced critique, students largely began with the approach or misperception that the foundational knowledge base in biology is infallible and is meant to be memorised and added to. Through experiences such as lab group discussions, students began to see that each 'fact' an article puts forth is actually an assertion that is both part of, and the end of, an argument. This change may be seen as an example defined by Carey (1999) as a conceptual transformation from seeing a concept defined by properties into one defined by relationships. In the case of this study, students came to understand that the outcome of a literature article is not a fact found by an experiment. Instead, the outcome is intrinsically connected to the choices made in the design of an experiment and its implications for the broader field. The emergence of primary literature as a threshold concept is consistent with findings from other studies (e.g. Urquhart et al., 2016; Timmerman et al., 2013) and endorses Kiley and Wisker's (2009) concept of *Framework*.

The second potential threshold concept found in our study resides within the broader concept of experimental design – specifically, a focus on controls within cellular and molecular biology experiments. Similar to primary literature, students also realise that controls and design are part of an argument within an experiment as opposed to just the properties of an experiment. This appears to happen as students themselves learn the limits of critical investigative environments with finite resources. Students must understand the purpose of each factor chosen for an experiment as they are forced by limited time or money to find more efficient methods. This push for simplicity in design also appears to happen with the need to interpret results. Flaws or excessive complexity in experimental design and/or the use of improper controls leads to results which are difficult to interpret. This need to clearly state what was found and why requires a more in-depth understanding of the relationships between design elements in experiments, as opposed to just seeing results as a property of the experiment. Such framing loosely falls under Kiley and Wisker's concept of *Argument* in that appropriate design contributes to a 'line of argument that [a] student could support with … their own findings' (p. 435). However, as noted elsewhere, this concept may be more serviceable if it were decomposed into more refined, better differentiated, and disciplinarily specific threshold concepts (Timmerman et al., 2013)

Timmerman and colleagues' (2013) previous elaboration on primary literature and experimental design asserted that mastery of these concepts to a certain level might be necessary precursors to advanced disciplinary understanding of hypotheses, study limitations, and robust conclusions. Although we did not find a consistent pattern pertaining to the timing of students' transformed understandings in these areas, participants spoke of meaningful periods of time before and after their attained understandings of these essential threshold concepts.

Students we interviewed reported developing these research skills neither synchronously nor independently. Respondents often expressed connections between learning to critique the primary literature they read – often in journal clubs – and concurrently learning to critique their own experiments. As students progressed, their treatment of the literature transitioned from perceiving it as infallible, to overly flawed, to a more balanced view. They similarly learned that their own experiments were neither worthless nor perfect but instead the best argument that could be made given the time, resources, and knowledge. However, participants were generally unable to articulate the mechanisms by which their experiences enabled them to cross thresholds. Many described briefly 'sticking with it' and 'riding it out,' while others simply said they 'didn't know' or 'couldn't explain it' because at some point, 'it just started making sense.'

Our discipline-specific findings also fit within previous research by Kiley and Wisker (2009) of more generic graduate level findings and add further evidence for how these occur. Specifically, we found evidence for two of their proposed threshold concepts: *Argument* and *Framework*. Their focus of an argument, which must be supported by defensible evidence, is a key feature of our findings on both primary literature as well as experimental design in biology. Although these concepts are distinct instances of argument in biology, they can also be seen as connected or as part of the web of connections between concepts (Davies & Mangan, 2005, 2007). Whereas primary literature focuses on arguments made by others in the field, experimental design focuses on the argument the individual is going to contribute to the field. Our division of argument into these components offers a better understanding of student learning within biology doctoral training. Finally, primary literature also fits within Kiley & Wisker's proposed threshold concept of framework which we found evidence for although students spoke of this in less detail.

Limitations

Given the exploratory and qualitative nature of this study, the findings are intended to provide more nuance and disciplinary context to prior work on threshold concepts in doctoral education. The sample was chosen deliberately to maximise the variation in institutional training experiences while constraining the population to a specific subfield and a research-intensive training context. The perspectives and experiences shared by participants in the current study can inform further investigation into threshold concepts as part of the graduate research training experience. However, they but should not be construed as inherently generalisable to either the larger population of cellular and molecular biology students, other branches of biology, or other STEM disciplines.

Implications

Much of the research on threshold concepts focuses on, and is limited to, theoretically proposed concepts by experts in the field. Threshold concepts are often identified by university faculty who draw upon their own instructional experiences to identify concepts that students struggle with (e.g. Kiley & Wisker, 2009). In other instances, these concepts are identified on the basis of items or concepts on which students test poorly (Shanahan, Foster, & Meyer, 2006). These approaches help to triangulate possible threshold concepts. However, they can also be limited by the perspective of faculty experiences, as they have 'long traversed the threshold' of specific concepts (Barradell, 2013, p. 243). Because of the irreversible nature of threshold concepts, we also need to compare these propositions to what students experience. What experts believe are threshold concepts or challenge students may not be the concepts that students find difficult or exist within frameworks that students find meaningful (Walker, 2013). The findings presented in this study represent the personal experiences of both experts who have passed these thresholds and current students at various stages of their doctoral training who may not have. As such, this approach adds to the diversity of available evidence regarding the nature of doctoral students' intellectual development within the discipline of the biological sciences.

A further benefit of our focus on the personal experiences of doctoral students in the biological sciences is an increased understanding of what knowledge was troublesome in these programmes and why. A better understanding of initial thoughts and perspectives of biological research throughout their programmes may inform how the development and integration of this new knowledge might necessarily transform previous understanding.

Our results hold several implications for the development of doctoral programmes in the biological sciences. Understanding the difficulties of learning to appropriately evaluate primary literature and design experiments, and their potential functioning as respective threshold concepts, point to a need for deliberate focus on them within doctoral biology training – especially at the outset. Such an understanding would allow mentors to better provide explicit guidance through the integration of concepts that some students may be unable to navigate independently. Because learning these concepts may lead to both uncertain liminal states for students (Baillie, Bowden, & Meyer, 2013; Meyer & Land, 2003) and can prove troublesome, students' performance may manifest as simple mimicry without deeper understanding (Walker, 2013). In more extreme cases, failure to acquire key threshold concepts has also been associated with dropping out of a degree programme altogether (Land, Cousin, Meyer, & Davies, 2005). It is imperative that we understand whether students have successfully integrated these concepts into their understanding.

The second implication for biology graduate training is the need to prioritise certain concepts and order them for efficiency. If student understanding of primary literature and experimental design must come before an understanding of hypotheses or implications as suggested by Timmerman et al. (2013) then the structure of research training must reflect this. It would not only be inefficient to build understanding of one concept if its necessary precursors are missing, but may decrease student retention if the order of presentation makes the integration of them unnecessarily difficult.

Finally, our findings point not just to areas where students had difficulty with potential threshold concepts but also to what experiences helped them eventually integrate these ideas. There is considerable evidence that mentors both misperceive student ability (Hinds, 1999; Feldon, Maher, Hurst, & Timmerman, 2015; Nickerson, 1999) as well progress (Hinds, Patterson, & Pfeffer, 2001; Wittwer, Nückles, & Renkl, 2008). Therefore, an understanding of what worked from a student perspective may be a more effective way of determining what experiences are most supportive of successful student integration of threshold concepts. For example, most students spoke of journal clubs as being the most important way they learned to critique primary literature and experimental designs. Through seeing how peers, postdoctoral students, and professors were able to find both faults as well as strengths in both research articles and their lab's experimental findings, students were able to apply these ideas to their own reading and research. A better understanding and consensus of evidence around these experiences can help inform the biology research training experience and place the appropriate interventions at the optimal time to ensure successful progress through these programmes.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by Directorate for Education and Human Resources [grant numbers 1242369, 1431234].

Note

1. To protect the confidentiality of the participating institutions, faculty, and students, institution names and specific locations are withheld.

References

Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., Roberts, K., & Walter, P. (2014). *Molecular biology of the cell* (6th ed.). New York, NY: Garland Science.

Alberts, B., Kirschner, M. W., Tilghman, S., & Varmus, H. (2014). Rescuing US biomedical research from its systemic flaws. *Proceedings of the National Academy of Sciences,111*, 5773–5777.

Austin, A. E. (2002). Preparing the next generation of faculty: Graduate school as socialization to the academic career. *The Journal of Higher Education*, *73*(1), 94–122. doi:10.1353/jhe.2002.0001

Austin, A. E., & McDaniels, M. (2006). Preparing the professoriate of the future: Graduate student socialization for faculty roles. In J. C. Smart (Ed.), *Higher education: Handbook of theory and research* (pp. 397–456). Dordrecht,The Netherlands: Springer.

Baillie, C., Bowden, J. A., & Meyer, J. H. F. (2013). Threshold capabilities: Threshold concepts and knowledge capability linked through variation theory. *Higher Education*, *65*, 227–246. doi:10.1007/s10734-012-9540-5

Barradell, S. (2013). The identification of threshold concepts: A review of theoretical complexities and methodological challenges. *Higher Education*, *65*(2), 265–276. doi:10.1007/s10734-012-9542-3

Boyer, R. (2003). Concepts and skills in the biochemistry/molecular biology lab. *Biochemistry and Molecular Biology Education*, *31*(2), 102–105. doi:10.1002/bmb.2003.494031020192

Brewer, C. A., & Smith, D. (2011). *Vision and change in undergraduate biology education: A call to action*. Washington, DC: American Association for the Advancement of Science.

Caldwell, B., Rohlman, C., & Benore-Parsons, M. (2004). A curriculum skills matrix for development and assessment of undergraduate biochemistry and molecular biology laboratory programs. *Biochemistry and Molecular Biology Education*, *32*(1), 11–16. doi:10.1002/bmb.2004.494032010295

Carey, S. (1999). Knowledge acquisition: Enrichment or conceptual change. In E. Margolis, & S. Laurence (Eds.), *Concepts: Core readings* (pp. 459–487). Boston,MA: MIT Press.

Chubin, D. (1983). *Sociology of sciences: An annotated bibliography on invisible colleges, 1972–1981*. New York, NY: Garland.

Davies, P., & Mangan, J. (2005). Recognising threshold concepts: An exploration of different approaches. In *European association in learning and instruction conference (EARLI)* (Vol. 23). Nicosia, Cyprus.

Davies, P., & Mangan, J. (2007). Threshold concepts and the integration of understanding in economics. *Studies in Higher Education*, *32*, 711–726. doi:10.1080/03075070701685148

Delamont, S., & Atkinson, P. (2001). Doctoring uncertainty: Mastering craft knowledge. *Social Studies of Science*, *31*(1), 87–107. doi:10.1177/030631201031001005

Erickson, F. (1986). Chapter 5: Qualitative methods in research on teaching. In M. C. Wittrock (Ed.), *Handbook of research on teaching (3rd ed.)* (pp. 119–161). New York,NY: Macmillan.

Feldon, D. F. (2010). Do psychology researchers tell it like it is? A microgenetic analysis of research strategies and self-report accuracy. *Instructional Science*, *38*, 395–415. doi:10.1007/s11251-008-9085-2

Feldon, D. F., Maher, M. A., Hurst, M., & Timmerman, B. (2015). Faculty mentors', graduate students', and performance-based assessments of students' research skill development. *American Educational Research Journal*, *52*, 334–370. doi:10.3102/0002831214549449

Feldon, D. F., Maher, M., & Timmerman, B. (2010). Performance-based data in the study of STEM graduate education. *Science*, *329*, 282–283. doi:10.1126/science.1191269

Feldon, D. F., Peugh, J., Timmerman, B. E., Maher, M. A., Hurst, M., Strickland, D., … Stiegelmeyer, C. (2011). Graduate students' teaching experiences improve their methodological research skills. *Science*, *333*(6045), 1037–1039. doi:10.1126/science.1204109

Feldon, D. F., Shukla, K., & Maher, M. A. (2016). Faculty-student coauthorship as a means to enhance STEM graduate students' research skills. *International Journal for Researcher Development*, *7*, 178–191. doi:10.1108/IJRD-10-2015-0027

Ferreira, M. (2003). Gender issues related to graduate student attrition in two science departments. *International Journal of Science Education*, *25*, 969–989. doi:10.1080/09500690305026

Gilbert, R., Balatti, J., Turner, P., & Whitehouse, H. (2004). The generic skills debate in research higher degrees. *Higher Education Research & Development*, *23*(3), 375–388. doi:10.1080/0729436042000235454

Golde, C. M., Conklin Bueschel, A., Jones, L., & Walker, G. E. (2009). Advocating apprenticeship and intellectual community: Lessons from the Carnegie Initiative on the doctorate. In R. G. Ehrenberg & C. V. Kuh (Eds.), *Doctoral education and faculty of the future* (pp. 53–64). Ithaca, NY: Cornell University Press.

Golde, C. M., & Dore, T. M. (2001). *At cross purposes: What the experiences of today's doctoral students reveal about doctoral education* (pp. 63). Philadelphia,PA: Pew Charitable Trusts.

Gross, A. J., & Mantel, N. (1967). The effective use of both positive and negative controls in screening experiments. *Biometrics*, *23*, 285–295. doi:10.2307/2528162

Hinds, P. J. (1999). The curse of expertise: The effects of expertise and debiasing methods on predictions of novice performance. *Journal of Experimental Psychology: Applied*, *5*, 205–221. doi:10.1037/1076-898X.5.2.205

Hinds, P., Patterson, M., & Pfeffer, J. (2001). Bothered by abstraction: The effect of expertise on knowledge transfer and subsequent novice performance. *Journal of Applied Psychology*, *86*, 1232–1243. doi:10.1037/0021-9010.86.6.1232

Isaac, P. D., Quinlan, S. V., & Walker, M. M. (1992). Faculty perceptions of the doctoral dissertation. *The Journal of Higher Education*, *63*(3), 241–268. doi:10.2307/1982014

Keefer, J. M. (2015). Experiencing doctoral liminality as a conceptual threshold and how supervisors can use it. *Innovations in Education and Teaching International*, *52*, 17–28. doi:10.1080/14703297.2014.981839

Kiley, M. (2009). Identifying threshold concepts and proposing strategies to support doctoral candidates. *Innovations in Education and Teaching International*, *46*(3), 293–304. doi:10.1080/14703290903069001

Kiley, M. (2015). 'I didn't have a clue what they were talking about': PhD candidates and theory. *Innovations in Education and Teaching International*, *52*, 52–63. doi:10.1080/14703297.2014.981835

Kiley, M., & Wisker, G. (2009). Threshold concepts in research education and evidence of threshold crossing. *Higher Education Research & Development*, *28*(4), 431–441. doi:10.1080/07294360903067930

Kleinman, D. L. (1998). Untangling context: Understanding a university laboratory in the commercial world. *Science, Technology, & Human Values*, *23*, 285–314. doi:10.1177/016224399802300302

Knorr-Cetina, K. (1997). What scientists do. In T. Ibáñez, & L. Íñiguez (Eds.), *Critical social psychology* (pp. 260–272). London: Sage.

Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press.

Labov, J. B., Reid, A. H., & Yamamoto, K. R. (2010). Integrated biology and undergraduate science education: A new biology education for the twenty-first century? *CBE – Life Sciences Education*, *9*(1), 10–16. doi:10.1187/cbe.09-12-0092

Land, R., Cousin, G., Meyer, J. H., & Davies, P. (2005). Threshold concepts and troublesome knowledge (3): implications for course design and evaluation. In C. Rust (Ed.), *Improving student learning diversity and inclusivity* (pp. 53–64). Oxford: Oxford Centre for Staff and Learning Development.

Lipschutz, S. S. (1993). Enhancing success in doctoral education: From policy to practice. *New Directions for Institutional Research*, *1993*(80), 69–80. doi:10.1002/ir.37019938008

Lovitts, B. (2001). *Leaving the ivory tower: The causes and consequences of departure from doctoral study*. Lanham, MD: Rowman & Littlefield.

Lovitts, B. E. (2008). The transition to independent research: Who makes it, who doesn't, and why. *The Journal of Higher Education*, *79*(3), 296–325. doi:10.1353/jhe.0.0006

Maher, M. A., Timmerman, B. E., Feldon, D. F., & Strickland, D. (2013). Factors affecting the occurrence of faculty-doctoral student coauthorship. *The Journal of Higher Education*, *84*, 121–143. doi:10.1080/00221546.2013.11777280

Mahmood, T., & Yang, P. (2012). Western blot: Technique, theory, and trouble shooting. *North American Journal of Medical Sciences*, *4*, 429–434. doi:10.4103/1947-2714.94940

Meyer, J. (2008). *Threshold concepts within the disciplines*. Rotterdam: Sense.

Meyer, J. H. F., & Land, R. (2003). Threshold concepts and troublesome knowledge (1): Linkages to ways of thinking and practising with the disciplines. In C. Rust (Ed.), *Improving student learning – Ten years on* (pp. 1–16). Oxford: OCSLD.

National Center for Science and Engineering Statistics. (2017). *Survey of graduate students and post-doctorates in science and engineering*. Washington, DC: National Science Foundation. Retrieved from https://www.nsf.gov/statistics/srvygradpostdoc/

Nickerson, R. S. (1999). How we know – And sometimes misjudge – What others know: Imputing one's own knowledge to others. *Psychological Bulletin*, *125*(6), 737–759. doi:10.1037/0033-2909.125.6.737

Paglis, L. L., Green, S. G., & Bauer, T. N. (2006). Does adviser mentoring add value? A longitudinal study of mentoring and doctoral student outcomes. *Research in Higher Education*, *47*(4), 451–476. doi:10.1007/s11162-005-9003-2

Pole, C. (2000). Technicians and scholars in pursuit of the PhD: Some reflections on doctoral study. *Research Papers in Education*, *15*, 95–111. doi:10.1080/026715200362961

Price, D. (1965). Networks of scientific papers. *Science*, *149*, 510–515. doi:10.1126/science.149.3683.510

Ruxton, G., & Colegrave, N. (2010). *Experimental design for the life sciences* (3rd ed.). New York, NY: Oxford University Press.

Schraagen, J. M. (1993). How experts solve a novel problem in experimental design. *Cognitive Science*, *17*(2), 285–309. doi:10.1016/0364-0213(93)90013-X

Schunn, C. D., & Anderson, J. R. (1999). The generality/specificity of expertise in scientific reasoning. *Cognitive Science*, *23*(3), 337–370. doi:10.1016/S0364-0213(99)00006-3

Shanahan, M., Foster, G., & Meyer, J. H. F. (2006). Operationalising a threshold concept in economics: A pilot study using multiple choice questions on opportunity cost. *International Review of Economic Education*, *5*(2), 29–57. doi:10.1016/S1477-3880(15)30119-5

Timmerman, B., Feldon, D. F., Maher, M., Strickland, D., & Gilmore, J. A. (2013). Performance-based assessment of graduate student research skills: Timing, trajectory, and potential thresholds. *Studies in Higher Education*, *38*, 693–710. doi:10.1080/03075079.2011.590971

Urquhart, S., Maher, M. A., Feldon, D. F., & Gilmore, J. (2016). Factors associated with novice graduate student researchers' engagement with primary literature. *International Journal for Researcher Development*, *7*, 141–158. doi:10.1108/IJRD-11-2015-0029

Walker, G. (2013). A cognitive approach to threshold concepts. *Higher Education*, *65*(2), 247–263. doi:10.1007/s10734-012-9541-4

Weidman, J. C., Twale, D. J., Stein, E. L., & Leahy, E. (2001). *Socialization of graduate and professional students in higher education: A perilous passage?* (pp. 1–139). Washington, DC: Office of Educational Research and Improvement.

Wittwer, J., Nückles, M., & Renkl, A. (2008). Is underestimation less detrimental than overestimation? The impacts of experts' beliefs about a layperson's knowledge on learning and question asking. *Instructional Science*, *36*, 27–52. doi:10.1007/s11251-007-9021-x