Faculty Mentors', Graduate Students', and Performance-Based Assessments of Students' Research Skill Development

David F. Feldon, Michelle A. Maher, Melissa Hurst and Briana Timmerman

The online version of this article can be found at:
http://aer.sagepub.com/content/early/2014/09/25/0002831214549449

Published on behalf of

AERA AMERICAN EDUCATIONAL RESEARCH ASSOCIATION

American Educational Research Association

and
§SAGE

http://www.sagepublications.com

Additional services and information for *American Educational Research Journal* can be found at:

Email Alerts: http://aerj.aera.net/alerts

Subscriptions: http://aerj.aera.net/subscriptions

Reprints: http://www.aera.net/reprints

Permissions: http://www.aera.net/permissions

>> OnlineFirst Version of Record - Sep 25, 2014

What is This?

Faculty Mentors', Graduate Students', and Performance-Based Assessments of Students' Research Skill Development

David F. Feldon
Utah State University
Michelle A. Maher
University of South Carolina
Melissa Hurst
University of Virginia
Briana Timmerman
University of South Carolina

Faculty mentorship is thought to be a linchpin of graduate education in STEM disciplines. This mixed-method study investigates agreement between student mentees' and their faculty mentors' perceptions of the students' developing research knowledge and skills in STEM. We also compare both assessments against independent ratings of the students' written research proposals. In most cases, students and their mentors identified divergent strengths and weaknesses. However, when mentor-mentee pairs did identify the same characteristics, mentors and mentees disagreed about the mentee's abilities in 44% of cases in the Fall semester and 75% of cases in the Spring semester. When compared against performance-based assessments of mentees' work, neither faculty mentors' nor their mentees' perceptions aligned with rubric scores at rates greater than chance in most categories.

KEYWORDS: graduate education, assessment, self-assessment, research skills

Introduction

Across the disciplines, faculty mentorship is considered essential in the development of graduate students' research skills (Barnes & Austin, 2009; Chan, 2006; Clark, Harden, & Johnson, 2000; Paglis, Green, & Bauer, 2006; Walker, Golde, Jones, Conklin Bueschel, & Hutchings, 2008). Such mentorship includes "establish[ing] a working relationship with a student and shepherd[ing] her or him through the [degree] process to completion" (Nettles & Millett, 2006, p. 98) as well as identifying areas in which the

student may not be developing appropriately and providing the necessary support to close perceived gaps (Ahern & Manathunga, 2004).

Surprisingly, however, relatively few studies examine the assessment of graduate students' skill development. The studies that do exist typically are phenomenological in nature and address either the faculty mentors' perceptions or the student mentees' perceptions, but not both (e.g., Austin, 2002; Gardner, 2007, 2009; Golde & Dore, 2001; Lovitts, 2008; but see Denicolo, 2003). Consequently, there are few data that can inform our understanding of the degree of alignment between the perspectives of mentors and mentees on mentees' skills. Given the frequency with which graduate students' feelings of preparedness for conducting independent research (i.e., self-efficacy)—especially in STEM (science, technology, engineering, mathematics) disciplines—serve as proxies for the skills acquired during their training (e.g., Austin, 2002; Delamont & Atkinson, 2001; Golde & Dore, 2001; Lovitts, 2001; Pole, 2000), it is important to understand the extent to which that perspective may converge with those of their mentors.

The current study investigates agreement between student mentees' and their faculty mentors' perceptions of the students' developing research skills in STEM disciplines. Further, each of these two perspectives is compared against independently scored measures of student participants' performance on written research proposals to determine the triangulated alignment among these common sources of data regarding graduate students' skill development. STEM disciplines provide a unique (and perhaps even biased) context within which to consider this triangulated alignment. Unlike their counterparts in the humanities and social sciences, STEM faculty and students often collaborate on common problems of inquiry, and faculty are more likely to perceive students as belonging to a collective, not individualistic, research effort (Franke & Arvidsson, 2011; Parry, 2007).

DAVID F. FELDON is the director of the Utah State University STE²M Center and an associate professor in the department of Instructional Technology and Learning Sciences, 6720 Old Main Hill, Logan, UT 84322-6720; e-mail: *david.feldon@usu.edu*. His research interests include the development of expertise in STEM disciplines, cognitive load theory, and graduate research training.

MICHELLE A. MAHER is an associate professor of higher education in the department of Educational Leadership and Policies at the University of South Carolina. Her research interests include interdisciplinary graduate student development, assessment in higher education, and faculty development, with a specific focus on bullying in the American academic workforce.

MELISSA HURST is the director of professional development at the University of Virginia. She provides career and student development services designed to assist master's and PhD students with the transition into the workplace.

BRIANA TIMMERMAN is the director of the Office of Instructional Practices and Evaluation at the South Carolina Department of Education. Her research interests focus on issues of assessment and skill development in sciences at the postsecondary level.

Correspondingly, Nettles and Millett (2006) found that STEM faculty-student interactions occur at a higher rate than among humanities and social sciences counterparts. Thus, generally speaking, STEM faculty should have ample opportunities to observe and assess their students' development as researchers.

Conceptual Framework and Review of Literature

Graduate education is often conceptualized as a process of acculturating or socializing a student into their chosen discipline (Austin & McDaniels, 2006; Gardner, 2009; Golde & Walker, 2006) with the ultimate purpose of the student becoming a skilled and, at the doctoral level, autonomous researcher (Gardner, 2008; Lovitts, 2008). Models of graduate student socialization, such as that offered by Weidman, Twale, and Stein (2001), posit that this process occurs within both formal and informal contexts and through an array of agents, including "significant others . . . in a department, a laboratory, a disciplinary network, or a university and its resources" (Pearson & Brew, 2002, p. 141).

Primary among these socialization agents is the student's faculty mentor. The mentor is thought to be essential to students' acquisition of disciplinary knowledge and skills through engaging students in a learning and teaching relationship conceptualized as a cognitive apprenticeship (Delamont, Atkinson, & Parry, 2000; Walker et al., 2008). The term *cognitive apprenticeship* was originally proposed to describe an instructional paradigm for developing cognitive expertise in formal secondary school settings (Brown, Collins, & Duguid, 1989; Collins, Brown, & Holum, 1991). However, it is now regularly applied to graduate education mentor-mentee pedagogical interactions, to the point that Golde, Conklin Bueschel, Jones, and Walker (2009) state, "We would go so far as to call apprenticeship the signature pedagogy of doctoral education" (p. 54).

Within traditional understandings of a cognitive apprenticeship, an experienced instructor or "master" guides a novice student or "apprentice" in his or her development of conceptual and problem-solving knowledge (Collins et al., 1991). Key goals of cognitive apprenticeship are to make visible the cognitive processes of the master or expert as he or she supports, through modeling, scaffolding, and coaching, the novice student's development of expertise. Fundamental to this development is the student's ability to learn the cognitive and metacognitive processes that underlie domain expertise. As related to the complex tasks associated with STEM graduate student performance, Delamont and Atkinson (2001) suggest that successful students "master . . . tacit, indeterminate skills and knowledge, produce usable results, and become professional scientists [who] learn to write public accounts of their investigations which omit the uncertainties, contingencies, and personal craft skills" (p. 88).

Conceptualizations of traditional apprenticeships suggest a one-to-one relationship in which an individual master guides an individual apprentice. However, cognitive apprenticeship is a pedagogical paradigm predicated on the conceptualization of learning as socially situated (Collins et al., 1991). Apprentices are immersed in a learning environment in which most or all are engaged in performing and perfecting the targeted skills. In this environment, apprentices can learn from others of varying levels of expertise. Collins and colleagues (1991) refer to this environment as a "community of practice" in which "the participants actively communicate about and engage in the skills involved in expertise, where expertise is understood as the practice of solving problems and carrying out tasks in a domain" (p. 16). More recent scholars also advocate for these types of learning environments. For example, Golde and colleagues (2009) promote the idea of "cascade mentoring" in a research lab setting, in which "post-doctoral fellows mentor senior graduate students, senior graduate students mentor junior graduate students, and junior graduate students mentor undergraduates" (p. 57).

Cascading or multiple mentoring is intuitively appealing, especially in light of concerns about inadequate individual faculty mentoring, which can be "marked by neglect, abandonment, and indifference" (Johnson, Lee, & Green, 2000, p. 136). Additionally, as STEM graduate students commonly participate in supervised laboratory-based research teams (Cumming, 2009; Parry, 2007), the extent to which a range of teammates at varying expertise levels contribute to a student's research skill development should not be overlooked. However, commonly accepted conceptualizations of graduate students' research skill development suggest that multiple mentoring can supplement—but does not replace—a primary faculty mentor (Carnegie Initiative on the Doctorate [CID], 2001; Paglis et al., 2006; Pearson & Brew, 2002).

Productive graduate faculty-student mentoring relationships ensure that students receive explicit guidance to cultivate and clarify necessary skills while performing authentic academic tasks as preparation to address new scholarly challenges (Pearson & Brew, 2002). This type of relationship is particularly critical at the initial stages of graduate work, as many novice students do not clearly understand what graduate study entails or their mentors' expectations for them or their work (Golde & Dore, 2001; Pole, Sprokkereef, Burgess, & Lakin, 1997). However, expressions of concern about uneven or inadequate graduate research supervision throughout the doctoral process are fairly prevalent (e.g., Anderson & Swazey, 1998; Lovitts, 2001), to the point that some have wondered if they are endemic to the process of doctoral education itself (Johnson et al., 2000). Thus, unsurprisingly, the review of relevant literature presented in the following raises serious concerns about the nature of skill assessment and related guidance within the mentor-mentee relationship.

Concerns Within the Mentor-Mentee Relationship

For the faculty mentor to directly influence a student's research skill development, several assumptions about the faculty, the student, and the nature of the relationship must hold. First, faculty mentors must themselves be well versed in the target skills and must accompany their modeling, scaffolding, and coaching with articulation that makes their thinking visible to the mentee. They must be able to estimate the extent to which the mentee's current knowledge and skills align with the scope and difficulty presented by the targeted task. While one could reasonably expect that faculty members hold the requisite level of expertise in the targeted skills, less is certain about their ability to effectively articulate instructional guidance. Dunbar (2000) reports that participants in laboratory meetings are typically unable to accurately reconstruct the collective reasoning processes by which discoveries and advancements are made. Even when explanations are provided immediately after decision making in an experimental design task, expert faculty tend to omit approximately 70% of the information necessary to replicate their processes (Feldon, 2010). Thus, the conscious availability of salient, verbalizable information to provide during mentorship discussions may be substantially limited.

Further, laboratory studies consistently find that experts underestimate the necessary time and effort required by novices to complete tasks and overestimate the relevant knowledge that novices possess (Hinds, 1999; Nickerson, 1999). These incorrect estimates have been shown to limit the effectiveness of the guidance that experts offer in instructional roles (Hinds, Patterson, & Pfeffer, 2001; Wittwer, Nückles, & Renkl, 2008).

Second, while the student must receive clear expectations regarding the development of research skills from his or her mentor, he or she must also assume responsibility for the development of those skills (CID, 2001; Kam, 1997). While the mentor makes visible the cognitive processes underlying his or her expertise, the student is encouraged to "externalize their own learning processes so they can gain access to and control over their own problem-solving strategies by articulating and reflecting on their knowledge, reasoning, or problem-solving processes" (Pearson & Brew, 2002, p. 140). However, while graduate students' research opportunities and other academic experiences are tailored to develop research skills as a prerequisite to becoming independent scholars, students often do not accurately perceive their own levels of skill (Dunning, Johnson, Ehrlinger, & Kriger, 2003; Ehrlinger & Dunning, 2003; Falchikov & Boud, 1989). This might be particularly true of early career graduate students who have yet to recognize, much less master, the tacit knowledge and skills that come with any given domain of expertise (Delamont & Atkinson, 2001).

Finally, the mentor-mentee relationship must be such that students and faculty jointly engage in critical reflection on research processes to facilitate

learning and foster the ability to transfer knowledge to new situations (Pearson & Brew, 2002). However, this joint reflection can prove difficult if a mentee's expectations for the mentor are misaligned. As Lagowski and Vick (1995) observe,

> The word *mentor* has several quite different accepted meanings: advisor, counselor, guide, preparer, monitor, teacher . . . faculty who think of mentoring almost exclusively in terms of imparting knowledge . . . will probably not be effective mentors for students who expect advice, counsel, and guidance [and] a friend. (p. 79)

The relationship can be further complicated given that mentors' and mentees' perceptions of mentees' research skill level often differ, with students rating their skill levels more highly than do faculty (Cox & Andriot, 2009; Kardash, 2000). Finally, even tangential aspects such as personality characteristics and types of interactions with the graduate student can impede accurate judgments of skill (Johnson, 2002).

Ability to Predict Performance

Although no studies have previously examined STEM faculty or graduate students, a number of them examine the ability of individuals to predict their performance and that of others. Generally, individuals' estimates of their abilities do not converge with the quality of their performance (Dunning et al., 2003). A meta-analysis by Stajkovic and Luthans (1998) found that across studies, self-efficacy beliefs predict no more than 25% of the variance in participants' actual performance during low-complexity, artificial tasks. For complex, authentic tasks such as those involved in scholarly research, self-efficacy beliefs only account for approximately 4% of the variation in performance on average.

Similarly, experts' estimates of students' abilities tend not to align with students' demonstrated performance (Hinds, 1999). Although this issue has not been examined with STEM faculty mentoring graduate students, studies in other domains have found that supervisors' opinions have very little power to predict student performance. For example, in medicine, supervising faculty with substantial time supervising their students in clinical circumstances fail to predict any significant variance in those students' performance on the National Board of Medical Examiners (NBME) clinical subject examination (Goldstein et al., 2014).

These findings collectively raise serious concerns about the nature of skill assessment and related guidance that occurs within the graduate mentor-mentee relationship. The current study investigates these issues by addressing the following three questions:

Research Question 1: To what extent do faculty mentors and their student mentees identify the same types of research skills and knowledge as aspects of the mentees' abilities?

Research Question 2: What is the level of agreement between mentors' and mentees' perceptions of the mentees' strengths and weaknesses in conducting research?

Research Question 3: What is the relationship between mentors' and mentees' perceptions and student mentees' performances on an authentic scholarly task?

Methods

The current study is part of a larger project examining the impacts of STEM graduate students' teaching and research experiences on the development of their research skills during the first years of their degree programs. We employ a mixed-methodological approach, using qualitative interview data from mentor-mentee pairs and double-blind quantitative assessments of student performance on sole-authored research proposals using a previously validated rubric (Feldon et al., 2011; Timmerman, Strickland, Johnson, & Payne, 2011).

This approach is unique in several ways. First, it draws on data from the mentor and the mentee independently to gain insight into the nature and influence of this relationship. Most knowledge about the graduate mentoring process is typically based on data collected exclusively from doctoral students without corresponding data from their mentors (Barnes & Austin, 2009). Second, while graduate advising is considered critical for students' skill development, few studies employ methods that permit direct assessment of students' acquired skills through performance-based measures (Feldon, Maher, & Timmerman, 2010). Third, the efforts reported here examine the early years of STEM graduate education, where faculty mentoring may arguably matter most for student success due to the importance of establishing a strong intellectual and cultural foundation in the degree program and research lab (Boyle & Boice, 1998; Golde, 1998).

Methodological Framework

This study utilizes a concurrent triangulation design in which both qualitative and quantitative data are collected and analyzed independently and then integrated for interpretation to assess the corroboration of inferences attained through multiple approaches (Creswell, Plano Clark, Gutmann, & Hanson, 2003; Greene, Caracelli, & Graham, 1989). To compare multiple, possibly divergent representations of the same phenomenon (student research skill), we adopt a phenomenographic perspective. This approach frames personal conceptions as relational objects between a reality external to the individual and the personal (phenomenological) meaning that is made

7

from it. Thus, its epistemic and logical assumptions acknowledge that there can be a true "fact of the matter" while avoiding the imposition of a binary *right* or *wrong* standard on the views offered from differing perspectives (Svensson, 1997). In this way, it is possible to meaningfully compare and contrast differing perspectives without privileging one over another. Further, phenomenography asserts differences in perspective linked to structural differences in social relationships, which aligns with the study's data drawn from students and their mentors.

Within this framework, we position three sources of conceptions for comparison: students' self-perceptions of their skill and knowledge pertinent to performing research tasks in their respective disciplines, faculty mentors' perceptions of their students' abilities, and raters' blinded, rubric-based assessments of students' written research proposals. Mentors' and mentees' perceptions were elicited through semi-structured interviews from which selected quotes were drawn. Broader categories were identified through analysis of these quotes. Thus, data are considered both within the context of the individual interview from which they came and within the context of the broader pool of meanings (Åkerlind, 2005; Marton & Pong, 2005).

Similarly, rubric-based ratings of the written research proposals apply a shared set of meanings, represented as criterion-based scores, within performance categories (e.g., experimental design, use of primary literature, etc.) and across levels of performance (i.e., quality scored on a 0–3 scale). The specific categories and criteria were developed iteratively by a team of STEM faculty and educational researchers and validated across disciplines (for details, see Timmerman, 2008; Timmerman et al., 2011). Differences across performance-level categories are reflected quantitatively using directly observable criteria assigned to each level of performance within each category (Johnson, Penny, & Gordon, 2009). These scores are interpreted within the context of the study by classifying them as above or below the sample mean on each rubric element to reflect relative strength or weakness on a normative basis within the context of the institutions where the study took place. Conceptions from all three sources are compared quantitatively to determine the frequency of alignment among these perspectives. More specific details are provided in the following sections.

Participants

Student participants (n = 81) were enrolled full-time in research-intensive master's and doctoral degree programs in STEM disciplines. Sixty-nine participants were enrolled in doctoral or master's level programs at a large research-intensive university in the Southeastern United States. The remaining 12 participants were enrolled in similarly intensive master's degree research programs at one of two smaller master's level public institutions in the Eastern United States. Participants from multiple institutions were

included to mitigate the potential influence of a single institution's culture on study findings. No differentiating trends in outcomes were detected as a function of institution, so affiliation data were not retained in the analyses.

All were volunteers participating in a larger study examining the development of teaching and research skills of STEM graduate students supported by the U.S. National Science Foundation (2007–2010). All signed a consent form that indicated all performance-based and interview data would remain confidential. For their participation in the larger study, student participants received a financial stipend. The sample was recruited from all STEM area programs, and students were retained for participation if they were actively engaged in supervised research and had not yet begun a dissertation or thesis project. Table 1 displays the program affiliation, year, and gender for each of the 81 students. No differentiating trends in outcomes were detected as a function of either these factors or degree program (MS vs. PhD), so they were not retained as independent variables in the analyses.[1] Race/ethnicity data were not reported by participants.

In addition to consenting to participate, all graduate students gave consent for researchers to conduct confidential interviews with their existing faculty mentors. To ensure that each faculty-student relationship did in fact adhere to the defining characteristics of a mentor-mentee relationship, we intentionally asked students to direct us to the faculty member they considered as their research mentor. In all cases used in this study, the student-identified faculty member was the student's primary research supervisor. Seventy-two faculty agreed to participate in the study after student permissions to contact them were received. Faculty participants were not compensated. Two faculty served as mentors for 2 student participants each, yielding a total of 74 intact mentor-mentee pairs whose relationships existed prior to the start of the study. The gender of the mentors interviewed was not considered a relevant factor, as published literature has not indicated that the quality of mentorship in STEM disciplines (Green & Bauer, 1995; Waldeck, Orrego, Plax, & Kearney, 1997) or mentoring behaviors/practices (Zhao, Golde, & McCormick, 2007) differ by same- or cross-gender mentor-mentee relationships. Further, the interview data did not suggest that participants found gender or race/ethnicity to be salient to the quality of their relationships with their mentors, as no participant addressed them in any way.

Based on interview data with participants, when admitted to their respective degree programs, students were generally matched with faculty mentors holding similar research interests as an outcome of the admissions process. However, some were preselected and/or recruited by individual faculty who perceived notable aptitude in a student in addition to similar interests. Several other students worked with faculty mentors with whom they shared a research interest after first spending some time working under the mentorship of a different faculty member. As described in Maher, Gilmore, Feldon, and Davis (2014), the vast majority of participants reported

9

Graduate Student Participants by Programmatic Affiliation and Year in Program

Discipline	First Year		Second Year		Third Year		Total
	Female	Male	Female	Male	Female	Male	
Engineering	**2**	**9**	**3**	**8**	**2**	**2**	**26**
Chemical engineering	0	0	1	2	0	2	5
Civil engineering	0	1	0	1	0	0	2
Computer science	0	2	0	1	1	0	4
Electrical engineering	1	1	0	1	0	0	3
Mechanical engineering	1	3	2	2	1	0	9
Nuclear engineering	0	2	0	1	0	0	3
Mathematics/statistics	**0**	**1**	**0**	**0**	**2**	**0**	**3**
Science	**13**	**16**	**10**	**8**	**3**	**2**	**52**
Biochemistry	1	0	0	0	0	0	1
Biology	3	3	0	2	1	0	9
Biomedical	2	2	1	0	0	0	5
Biotechnology	0	1	1	3	0	0	5
Chemistry	3	2	0	0	0	0	5
Environmental management	1	0	0	0	0	0	1
Epidemiology	1	1	0	0	0	0	2
Exercise science	2	0	0	0	0	1	3
Geosciences	0	1	0	1	2	0	4
Marine science	0	5	6	1	0	1	13
Microbiology	0	0	1	0	0	0	1
Molecular biology	0	1	1	1	0	0	3
Total	**15**	**26**	**13**	**16**	**7**	**4**	**81**

strong, positive relationships with their advisors, and all participants were able to describe interactions reflective of ongoing mentorship with the faculty members they identified as their mentors (e.g., coauthorship, discussions about the interpretation of data, development of research skills, etc.). No differentiating trends in interview data were detected as a function of how mentees were originally paired with their mentors, so these factors were not incorporated into current analyses. Additionally, no student participants were on laboratory or research group rotations and no participants changed advisors during the study, so mentor relationships were consistent across the fall and spring semesters.

Due to mentor scheduling constraints and participant attrition, not all mentors and mentees were interviewed in both fall and spring. All mentors were interviewed in the fall, and 50 of 72 were interviewed in the spring. A review of mentors' faculty ranks revealed a roughly even distribution across assistant, associate, and full professor ranks (27 assistant, 15 associate, and

10

25 full). Of the 81 mentees interviewed in fall, 69 were interviewed in spring. A total of 69 matched mentor-mentee pairs were interviewed in fall, and 49 matched pairs were interviewed in spring. In both fall and spring, data were retained for participants for whom the other member of the pair was not available as appropriate for specific analyses.

Data Collection

Relevant data were collected from graduate students and their respective faculty mentors during one academic year of the three-year project (2007–2008, 2008–2009, or 2009–2010). Student participants submitted written research proposals related to their academic focal areas in early fall. They then revised these proposals over the course of the academic year and resubmitted them in late spring. The research team provided no feedback to the participants between the fall and spring submissions, though participants were free to seek independent feedback from other sources typical of their academic support networks and programmatic resources.

Both students and their mentors participated independently in semi-structured interviews in early fall and late spring to discuss the students' areas of strength, weakness, and growth as developing researchers. We selected interview-based, unguided elicitations of perceived strengths and weaknesses over a Likert-style survey instrument of a discrete list of skill-specific questions to avoid the collection of ungrounded opinions generated in response to a presented query rather than a true assessment of skill. Research in survey methods indicates that many people will answer selected-response (i.e., forced-choice) queries despite a lack of knowledge or opinion on a topic. Further, providing a "don't know/no opinion" option fails to mitigate this tendency even when respondents are highly educated and surveys are not anonymous, as is the case with our sample (Bishop, Tuchfarber, & Oldendick, 1986; Krosnick et al., 2002; Schwarz, 1999). Due to the direct examination of convergence between advisors' and students' opinions and performance-based assessments, the analytic priority in this study is to maximize the validity of participant responses. The interview approach minimizes the likelihood that informants will unintentionally provide spurious data regarding students' research skills, though it does so at a cost to the quantity of data available for comparison against the perspective of the other member of the mentor-mentee dyad and the scores from participants' written research proposals.

Research Proposals

To gain an authentic, performance-based measure of research skills, student participants were asked to submit sole-authored research proposals for projects in their respective areas of interest early in the fall term of their participation year. They were also asked to revise and resubmit their proposals

in late spring of the same academic year. Participants were instructed to describe the relevant literature and design for their proposed research as well as anticipated results, other potential outcomes, and the significance of these in relation to their research questions or hypotheses. A summary of the key evaluation criteria was provided and explained to them during a participant orientation meeting.

During the orientation, it was explained to participants that they were free to select a topic of their choosing on which to write their research proposals. Specifically, they were encouraged to select a topic that would be directly related to their area of research interest and ideally be developed to contribute to required work, including conference and thesis proposals. Participants were also aware that although their proposals would be scored using a rubric that applied the criteria provided to them, the scores would not be reported back to them or to their mentors.

Most participants reported informally that they used their proposals for an additional purpose beyond the research study itself, such as for classwork, research lab, or conference proposal. This information was interpreted as a positive indicator of both legitimate effort invested in the task and ecological validity (i.e., ''conducted in settings that occur in the culture or subculture for other than research purposes . . . [including] place, time, roles, and activities'' [Bronfenbrenner, 1976, p. 7]).

Interviews

Most interviews with mentees and with their mentors, respectively, were approximately 30 minutes in duration, although they ranged between 20 and 60 minutes. During the interviews in early fall and again in late spring, participants were asked to describe their/their mentee's strengths, weaknesses, and overall identity as researchers. Participants received the following interview instructions:

> I'd like to talk with you about [yourself/student's name] as a teacher and as a researcher. I would like to tape our conversation so that I can transcribe it for more careful study. I will protect your anonymity and I will safeguard this tape and our conversation, so please feel free to share your thoughts about [your/student's name] teaching and research with me. What questions do you have for me before we begin?

The investigators used a semi-structured interview format to direct the focus of the discussion and allow participants the freedom to expand on their views of research skills and their own growth. Interview protocol items included questions that addressed four topics: (a) participants' goals, (b) participants' self-assessment of skills, (c) participants' experiences that may have influenced skill development, and (d) questions regarding participants'

relationships with their respective advisors. To enhance rapport, graduate research assistants conducted the interviews of mentees, and senior members of the research team conducted faculty mentor interviews. Although there cannot be any guarantee of full disclosure by any participant in any interview, interviewers worked to develop trust with participants, and the information disclosed in the interviews displayed such trust when describing personal hopes, doubts, and failures related to participants' graduate studies.

Mentors were asked to describe their participating students' research strengths and weaknesses. Key interview questions included "How would you describe yourself/your student as a researcher?" and "What are your/ your student's strengths and/or weaknesses as a researcher?" All interviews were tape-recorded and transcribed.

Data Analysis

Research Proposals

Research proposals were assessed using a modified form of a previously validated rubric for evaluating scientific research skills in written laboratory reports (Timmerman et al., 2011; Timmerman, Johnson, & Payne, 2007). This rubric was generated by review of relevant research literature on academic peer review criteria (Cicchetti, 1991; Marsh & Ball, 1989; Marsh & Bazeley, 1999; Petty, Fleming, & Fabrigar, 1999), concepts of current best practices for scientific reasoning or inquiry in STEM disciplines (Haaga, 1993; Halonen et al., 2003; Kelly & Takao, 2002; Tariq, Stefan, Butcher, & Heylings, 1998; Topping, Smith, Swanson, & Elliot, 2000; Willison & O'Regan, 2007), and recursive review and application by university STEM faculty to undergraduate laboratory courses. The only modification to the rubric for its use in the current study was a change of the wording to allow for predicted results rather than actual results and subsequent similar wording changes for discussing affordances and limitations. These wording modifications occurred through discussion and consensus within the rating team. The research skills assessed by the rubric were: setting context for a study, framing testable hypotheses, attention to validity and reliability of methods, experimental design, appropriate selection of data for analysis, data analysis, basing conclusions on data, identifying study limitations and significance, and effective use of primary literature. Specific criteria for each performance level are available in the online appendix.

Proposals were received from participants electronically, checked for plagiarism using SafeAssign software, and assigned to trained raters based on subject matter. Any submitted proposal with issues relating to improper or insufficient citation was returned to the participant for correction. All such communications occurred confidentially between the research team and the specific participant involved, and resources providing guidance on appropriate citation practices were provided. The majority of issues

detected reflected a lack of familiarity with normative academic writing practices, rather than intentional deception (for further details and discussion, see Gilmore, Strickland, Timmerman, Maher, & Feldon, 2010).

At least two raters with graduate degrees in relevant STEM disciplines scored both submissions of each proposal, and any discrepant scores were resolved by discussion until consensus was reached (cf. Johnson et al., 2009; Johnson, Penny, Gordon, Shumate, & Fisher, 2005). Pre-consensus measures of interrater reliability reflected intraclass correlation coefficients (ICCs) for individual rubric elements between 0.6 and 0.9. ICCs are considered equivalent to Cohen's kappa but are appropriately applied to continuous (rather than categorical) values (Bloch & Kraemer, 1989; Fleiss & Cohen, 1973).

Interviews

A preliminary set of research skills was expected to emerge that mirrored themes based on previously published literature from the fields of scientific reasoning and scientific writing within and across mentor/mentee roles and disciplines (e.g., Haaga, 1993; Halonen et al., 2003; Feldon, Timmerman, Stowe, & Showman, 2010; Topping et al., 2000; Zimmerman, 2000). Examples of these expected themes included various skills related to a sense of context for their research ("the big picture"), drawing valid inferences from data, and methodological/technical skills in experimental design and data analysis. Because no notable differences were reflected in the responses based on role (student/mentor), institution, or discipline, these distinctions were not used for full analyses.

Semi-structured interview data were collected twice from each student participant and from as many faculty participants as possible. To compile interview data, transcripts were analyzed using the constant comparison approach (Glaser, 1965). Two members of the research team continuously compared statements regarding mentee skills and knowledge both within and across mentor and mentee groups to identify emergent themes. Due to the open-ended nature of the interview questions (e.g., "How would you describe yourself as a researcher?") and to preserve authenticity and richness of participant data, the research team identified multiple types of skills, knowledge, and personality characteristics. Only responses coded as skills and knowledge were used for the current analysis to optimize congruence with the categories in the performance-based rubric and to avoid the inappropriate imposition of performance-related labels on traits associated with personality or demeanor (e.g., curiosity, diligence, etc.).

Sixteen codes reflecting student mentees' skills ($n = 12$) and knowledge ($n = 4$) were identified in both the mentor and mentee interviews and evaluated as indicating strength or weakness. Redundant codes and positive/negative characterizations of the same skills were consolidated (e.g.,

"analytic skills" and "data analysis" codes were merged into a single code; strengths and weaknesses were coded as "analytic skill–strength" and "analytic skill–weakness," respectively). Of the 16 codes, 9 were theoretically aligned with proposal rating categories, as described in the following, and were retained for further analysis. These 9 codes are listed and accompanied by illustrative quotes describing strengths and weaknesses in Table 2.

Student mentee skill and knowledge codes identified in both the mentor and mentee interviews were theoretically aligned with the proposal rating categories. Some codes aligned easily with one or more categories (e.g., the code Literature Review aligned easily with the category Introduction/Context and with the category Primary Literature Use). The alignment between other codes and categories was more nuanced (e.g., alignment between the code Big Picture and the categories Introduction/Context and Limitations/Significance). Seven codes (Communication, Computer Skills, Field Work, Operating Lab Equipment, Technique, Terminology, and Theoretical Framework) did not align with any proposal rating category. Care was taken to be as conservative as possible in determining alignment between codes and categories. Arguably, many of the codes and categories share some features. However, to limit potentially spurious alignment rates, only those that were self-evidently aligned were selected for analysis. Code and category alignments, denoted by an X, are displayed in Table 3. Operational definitions and evaluation criteria used in the rubric categories are provided in the online Appendix.

Next, the extent to which interview data aligned with performance data was assessed comparing rates of alignment and misalignment within categories using a nonparametric one-sample binomial test. Due to the small sample size within categories, this test was selected to avoid the assumption of a normal distribution and to obtain an exact p value.

Separate tables were maintained for data collected in fall and spring, respectively. This decision was made for two reasons. First, despite full participation by students in both fall and spring, not all faculty mentors were available for interviews in both fall and spring. Such missing data in an aggregated table would have resulted in a truncated data set due to listwise deletion associated with missing data during statistical analysis. Second, the passage of time over the course of the academic year could have impacted the perceptions and the extent of agreement between pairs of faculty and student participants. Pooling the data across time points would have obscured this potential trend.

Results

The results of the paired mentor and mentee response analyses conducted to identify alignment of perceived skill and knowledge strengths and weaknesses are presented first. These results are followed by the

15

Table 2
**Skills and Knowledge Codes Derived From Interview
Analysis With Illustrative Quotes**

Codes	Perceived Strength	Perceived Weakness
Analysis skills	I think she's quite analytically talented, certainly excellent with the computer, and she's in the top half of our PhD candidates, even higher than that in terms of her analytical skills.[a]	Well I think I'm okay. I think I have some improvements to make especially from the analytical point of view.[b]
Big picture	I think in ways that connect to the past, what direction they've actually helped contemporary science, the field that I'm studying in, and trying to see if those giants that actually brought the field to this particular point, saw things in this particular way, what can be learned from their endeavors to actually move society forward based upon those skills and actions.[b]	[My weakness] is . . . big concept kind of stuff. Sometimes your head's down the pipette. Sometimes . . . and I kind of have dual bosses. One's like "Do your work! Do your work." And the other guy is usually like "let's think about things." . . . So that would be my limitation.[b]
Conceptual knowledge	He understands the theory and concepts about the work he's doing.[a]	I think the biggest disadvantage for me is that I did not work in this field before, so I need to learn more about this background.[b]
Critical thinking	I've gotten really good at thinking critically and asking the right questions and if I see something that doesn't make sense I and try to figure out why it doesn't make sense and try to find the reasons why.[b]	I don't know how to put stuff together yet, and I need to be able to do that, and so I don't know if I will be able to critically look at something and think outside the box, that is a major problem with me.[b]
Data interpretation	She has good skills . . . looking at the data analytically and being able to—even though she's building her quantitative analytical skills, she's already demonstrated skills in being able to interpret what the results mean.[a]	I need somebody who can really do some sort of major data analysis . . . even they don't understand, I don't care. I can have students who don't understand what they are doing, really, but able to do it accurately, and then give it to me, and I can supply the understanding. I don't mind that. But, it's tough if somebody can't do either and so I really don't know where he's going to go.[a]

(continued)

Table 2 **(continued)**

Codes	Perceived Strength	Perceived Weakness
Defining problem	I'd say, the strength that he has is—and it is more advanced than I see in most students at this stage—is that he has a clear understanding that you've got to have a well-defined problem statement and research question from which you derive your methodology. Too often I encounter students who want to exercise the methodology and then go look for a problem to fit that, and he is really positing good research questions.[a]	Coming up with researchable questions, really in terms of defining the scope of the research. I tried to let her have a lot of input on the research area for her master's thesis and it finally came up that one day she said "tell me what to do" so I had to really scope out the objectives and the researchable questions for her.[a]
Literature review	I can find the literature and also translate and apply that work into the real laboratory skill.[b]	There have been instances where I start to solve a problem, tell somebody else about it and they go, "I think I heard of that before," and what do you know, it's already been done. Something I've spent my time trying to figure out, somebody's already done that. So I should have spent more time getting resources, looking on the Internet, that kind of stuff.[b]
Math/statistical skills	I'm good at math.[b]	Knowledge of stats I know I could definitely improve on.[b]
Research design	I think that was one of her strengths and really her ability I think to both conceive of and then follow through on an experimental design. I think she's very good at that. Other than that, she was fantastic in field work and really good at experimental design.[a]	Probably a lot of [weaknesses]. Setting up an experiment that works, that would be a good one. Because mine didn't really work that well. I had to take my plan B.[a]

[a]Indicates a mentor perception of strength or weakness.
[b]Indicates a mentee perception of strength or weakness.

comparison responses of faculty mentors' and student mentees' interview-based assessments to performance-based measures of their skills as assessed by multiple reviewers. Together, results represent a systematic investigation of multiple perspectives on graduate students' research skills and knowledge that represent the degree of convergence amongst these data sources.

Comparison of Code Alignment in Mentor-Mentee Paired Responses

Analysis of mentor-mentee interview code alignment reflects general parity between the total number of identified strengths and weaknesses in mentees' research skill and knowledge. In total, mentors offered 80

Table 3
Comparison of Proposal Rubric Categories to Mentor/Mentee Skill and Knowledge Codes From Interviews

Rubric Categories	Analysis Skills	Big Picture	Conceptual Knowledge	Critical Thinking	Data Interpretation	Defining Problem	Literature Review	Math/Statistical Skills	Research Design
Introduction/Context		X	X				X		
Testable Hypothesis	X			X		X			
Validity/Reliability	X			X					
Experiment Design	X			X					X
Data Selection				X		X			
Data Analysis	X			X	X	X		X	
Discussions/Conclusions	X	X		X	X	X			
Limitations/Significance	X			X					
Primary Literature			X				X		

Note. X denotes code and category alignments.

perceptions of mentee strengths (41) and weaknesses (39) in fall and 39 perceptions of strengths (20) and weaknesses (19) in spring. Mentees offered 88 total perceptions of their strengths (46) and weaknesses (42) in fall and 51 total perceptions in spring (strengths = 27, weaknesses = 24).

Most mentor and mentee responses did not overlap with regard to the strengths or weaknesses identified. It should be noted that the term *overlap* as utilized here does not entail agreement between mentor and mentee on whether the skill identified was a strength or weakness. Rather, these are instances where both members of the pair simply identified the same skill as relevant to the mentee's abilities. Rates of agreement on strength or weakness are reported in the following paragraph. In fall, only 16 of the 69 intact mentor-mentee pairs interviewed (23.2%; $p < .001$) identified the same type of research skill or knowledge as being either a notable strength or weakness of the mentee. Fourteen pairs overlapped in one skill or knowledge area, and 2 pairs overlapped twice, yielding a total of 18 instances. In spring, of the 49 intact mentor-mentee pairs interviewed, only 4 pairs (8.2%; $p < .001$) identified the same research skill or knowledge area (1 instance per pair). Overlapping assessments (those identified by both members of a mentor-mentee pair) are shown in Table 4, with indications of agreement presented out of the total number of overlaps per category.

In fall, of the 18 instances in which mentor-mentee pairs identified the same skill or knowledge category as a mentee's strength or weakness, mentors and mentees agreed on whether the identified skill/knowledge was a strength or weakness in only 10 instances (55.6%; $p = .82$). In spring, of the 4 instances in which mentor-mentee pairs identified the same skill or knowledge category as a mentee's strength or weakness, mentors and mentees agreed on whether the identified skill/knowledge was a strength or weakness in only 1 instance (25.0%; $p = .63$).

For example, in one case of congruence between a mentee and her mentor, the mentee indicated that one of her research strengths was finding and reviewing literature relevant to her research:

> I think finding literature I'm very good at because fire shrimp are not heavily researched and neither were the snails that I worked with in Bermuda. So those are two species that there's not much about. But you need to have a lot of literature so it took me a while to get really good at finding useful things that maybe it's not about the organism but it can be related. (Mentee)

Similarly, her mentor praised her literature review skills as a strength:

> She picked right up on it immediately and started getting the literature and things that were related to the project. So she does a really good job searching the literature. (Mentor)

19

Frequency of Agreement Between Mentor- and Mentee-Perceived Strengths and Weaknesses in Student Skill and Knowledge for Fall and Spring

	Fall		Spring	
	Raw	%	Raw	%
Analysis Skills	0/2	0	0/1	0
Big Picture	2/3	67	0/1	0
Conceptual Knowledge	2/3	67	0/1	0
Critical Thinking	0/0		0/0	
Data Interpretation	0/0		0/0	
Defining the Problem	0/0		0/0	
Literature Review	2/3	67	1/1	100
Math/Statistical Skills	2/2	100	0/0	
Research Design	2/5	40	0/0	
Total	10/18	56	1/4	25

In 8 of 18 pairs (44.4%) in fall and 3 of 4 pairs (75.0%) in spring, however, mentors and mentees disagreed about the mentee's strengths and weaknesses. For example, a student described his weakness as a researcher, coded as conceptual knowledge, by stating:

> I'm pretty good with theoretical ideas on evolution. I think I have a good grasp on how evolution works and populations within species within the greater context. I'm gradually getting more skilled in the computational side of bioinformatics but I still think that my knowledge base as the biology student is my strong set. (Mentee)

However, his advisor's comments countered this student's self-assessment:

> So far his research is greatly slowed and inhibited by the lack of background knowledge. He has a lot to catch up with and fill up those gaps because research in biology is a very complex thing. (Mentor)

In another misaligned case, a student stated, "Probably a lot of things [are weaknesses]. Setting up an experiment that works, that would be a good one. Because mine didn't really work that well. I had to take my plan B." In contrast, her advisor observed that,

> One of her strengths and really her ability I think to both conceive of and then follow through on an experimental design. I think she's very good at that. Other than that, she was fantastic in field work and really good at experimental design.

Comparison of Perceptions Against Performance

Mentor and mentee perceptions of mentees' strengths and weaknesses in their research skills and knowledge were cross-tabulated with elements of each mentee's rated proposal score. When compared to the sample population's mean scores, mentors' perceptions of mentees' research skills were not always accurate predictors of their mentees' demonstrated research skills and knowledge. During the fall, 71 evaluative comments by mentors of their mentees' research skills and knowledge were relevant to rubric criteria. Of these, none predicted performance on the rubric at levels significantly greater than chance, with aggregate rates ranging from 0% to 58.3% (see Table 5). Comments regarding mentees' data analysis skills predicted rubric-assessed performance at a rate significantly worse than chance ($p <$ .001). This misalignment reflected mentors' perceptions of strength that corresponded to below average performance and their perceptions of weakness that corresponded to above average performance on the written research proposals.

During the spring interviews, mentors commented on mentees' strengths and weaknesses in the categories that aligned with the rubric criteria 30 times with rates of aggregate alignment occurring between 0% and 83.3%. Despite the notably higher end of this range, comments did not predict performance on rubric criteria at levels significantly greater than chance, due at least in part to smaller sample sizes per category. For example, only one comment was made regarding a mentee's analysis skills, which is consistent with norm-referenced performance in five out of six matching criteria. As a single case, however, its representativeness cannot be determined. In categories with sample sizes comparable to data collected in the fall (i.e., Literature Review, Conceptual Knowledge, Big Picture), rates of alignment did not differ significantly ($p =$ .13, $p =$ .50, and $p =$ .50, respectively, using McNemar's nonparametric test of related samples) with aggregate rates between 50.0% and 61.1% (see Table 6).

Like their mentors, mentees' self-assessments were mostly inconsistent with their rubric-assessed performance. In fall, mentees articulated perceptions of their research skills and knowledge strengths and weaknesses 90 times in categories that aligned with rubric criteria. Aggregate rates of alignment ranged between 7.1% and 87.5% (see Table 7). Three categories yielded rates of alignment that differed significantly from chance. Self-assessments of literature review skills accurately predicted performance as assessed by the rubric at an overall rate of 67.3% with 20 out of 26 cases for the introduction/context criterion ($p =$ .011). Self-assessments of mathematical/statistical skills aligned at a rate of 20% (*ns*). Mentees' assessments of their abilities to grasp the "big picture" aligned at an aggregate rate of 30.0%, with alignment between big picture and identifying study limitations and significance occurring in only 3 out of 15 cases ($p =$ .035).

Table 5
Proportion of Agreement Between Mentor-Perceived Strengths and Weaknesses in Student Skill and Knowledge and Rubric-Assessed Mentee Performance for Fall

Rubric categories	Analysis Skills	Big Picture	Conceptual Knowledge	Critical Thinking	Data Interpretation	Defining Problem	Literature Review	Math/Stat Skills	Research Design
Introduction/Context		3/6	5/11				7/14		
Testable Hypothesis	3/10			2/4		2/3			
Validity/Reliability	3/10			3/4					
Experiment Design	5/10			3/4					8/15
Data Selection				2/4		2/3			
Data Analysis	5/10			1/4	1/2	2/3		0/6*	
Discussion/Conclusion	6/10			2/4	1/2	1/3			
Limitations/Significance	4/10	2/6		2/4					
Primary Literature			5/11				8/14		
Mean percentage congruent	43.3	41.7	45.5	53.6	50.0	58.3	53.6	0.0	53.3

*$p < .05$ per binomial test.

Table 6

Proportion of Agreement Between Mentor-Perceived Strengths and Weaknesses in Student Skill and Knowledge and Rubric-Assessed Mentee Performance for Spring

	Analysis Skills	Big Picture	Conceptual Knowledge	Critical Thinking	Data Interpretation	Defining Problem	Literature Review	Math/Stat Skills	Research Design
Rubric categories									
Introduction/Context		5/9	3/7			0/0	3/7		
Testable Hypothesis	1/1			0/0					
Validity/Reliability	1/1			0/0					
Experiment Design	0/1			0/0					0/0
Data Selection				0/0		0/0			
Data Analysis	1/1			0/0	0/0	0/0		0/6*	
Discussion/Conclusion	1/1			0/0	0/0	0/0			
Limitations/Significance	1/1	6/9		0/0					
Primary Literature			5/7				4/7		
Mean percentage congruent	83.3	61.1	57.1	N/A	N/A	N/A	50.0	0.0*	N/A

*$p < .05$ per binomial test.

As with the mentors, mentees' spring interviews yielded fewer perceptions of research skills and knowledge that aligned with rubric criteria (29 statements vs. 90 in fall). Rates of alignment with rubric-assessed performance reflected a narrow range between 50.0% and 58.3%. No categories predicted performance at a level significantly different from chance (see Table 8).

Discussion of Results

Faculty mentors are thought to be pivotal influences in the scholarly and professional development of graduate students (e.g., Barnes & Austin, 2009; CID, 2001; Pearson & Brew, 2002; Walker et al., 2008). As eloquently described by Hawley (1993), they "do more than simply stand and point the way. They accompany their protégés through the entire process [of graduate study]" (p. 53). From this perspective, it is easy to understand why the intensity, frequency, and focus of mentor-mentee interactions leads to the assumption that mentors hold privileged insight into their students' abilities. However, these findings suggest that mentors' perceptions of mentees' research skill and knowledge rarely match those of their mentees. Even more disconcerting, mentors' perceptions fail to predict mentees' performance at levels better than chance.

Faculty Perceptions

Why would mentors' perceptions so sharply misalign with their mentees' own perceptions and demonstration of research skills and knowledge? Several explanations might be considered. First, because disciplinary knowledge develops slowly, it is possible that students in their first few years of graduate training may not be at a point where their knowledge acquisition—or lack thereof—captures their mentors' attention. The pace of faculty work has noticeably increased (Austin & McDaniels, 2006), and faculty time is a limited resource (Remler & Pema, 2009). In this context, faculty mentors make frequent decisions about both how and with whom to spend their time (Milem, Berger, & Dey, 2000). Faculty mentors may prioritize their time and attention to aid those students who are farther along in their program and therefore close to obtaining their degrees. In these cases, mentors may intentionally or unintentionally leave beginning students' skill development in the hands of more senior students and postdoctoral fellows. Thus, much early mentee skill development may occur while the mentee is not under the watchful eye of the mentor (Delamont & Atkinson, 2001; Pearson & Brew, 2002).

Second, it is also possible that the many demands on faculty members' time lead them to rely on abstract impressions of individual students. Consistent evidence from research in social and cognitive psychology suggests that cognitive busyness increases the likelihood that individuals rely

Table 7

Proportion of Agreement Between Mentee-Perceived Strengths and Weaknesses in Student Skill and Knowledge and Rubric-Assessed Mentee Performance for Fall

	Analysis Skills	Big Picture	Conceptual Knowledge	Critical Thinking	Data Interpretation	Defining Problem	Literature Review	Math/Stat Skills	Research Design
Rubric categories									
Introduction/Context		6/15	7/12				20/26*		
Testable Hypothesis	5/9			0/2		2/2			
Validity/Reliability	3/9			0/2					
Experiment Design	3/9			0/2					4/11
Data Selection				1/2		2/2			
Data Analysis	3/9			0/2	3/8	1/2		1/5	
Discussion/Conclusion	4/9			0/2	5/8	2/2			
Limitations/Significance	3/9	3/15*		0/2					
Primary Literature			6/12				15/26		
Mean percentage congruent	38.9	30.0	54.2	7.1	50.0	87.5	67.3	20.0	36.4

*$p < .05$ per binomial test.

Table 8
Proportion of Agreement Between Mentee-Perceived Strengths and Weaknesses in Student Skill and Knowledge and Rubric-Assessed Mentee Performance for Spring

	Analysis Skills	Big Picture	Conceptual Knowledge	Critical Thinking	Data Interpretation	Defining Problem	Literature Review	Math/Stat-Skills	Research Design
Rubric categories									
Introduction/Context		1/2	3/6				4/8		
Testable Hypothesis	0/0			0/0		0/0			
Validity/Reliability	0/0			0/0					
Experiment Design	0/0			0/0					0/0
Data Selection				0/0		0/0			
Data Analysis	0/0			0/0	2/7	0/0		3/6	
Discussion/Conclusion	0/0			0/0	5/7	0/0			
Limitations/Significance	0/0	1/2		0/0		0/0			
Primary Literature			4/6				4/8		
Mean percentage congruent	N/A	50.0	58.3	N/A	50.0	N/A	50.0	50.0	N/A

*p < .05 per binomial test.

on mental heuristics and generalized, stereotypic impressions (Feldon, 2007; Greenwald & Banaji, 1995; Macrae, Bodenhausen, Milne, & Wheeler, 1996; Sloman, 2002). Descriptions of general disposition might serve as a heuristic, quickly bringing to mind a particular mentee with a single adjective. Broad perceptions of a mentee as "unfocused," "aggressive," or "hard working" possibly divert attention from the actual skills or knowledge either demonstrated or in need of development. While the literature has not yet directly examined this topic, it may also be the case that mentors tend to spend less time with mentees perceived as less skilled or requiring more guidance in favor of those perceived to require less effort or guidance (but for suggestive evidence, see Green & Bauer, 1995).

Third, research practices in STEM fields may also contribute to this phenomenon and may have been doing so for some time (e.g., Merton, 1973). In a world of multi-teamed research using multidisciplinary perspectives, the research process is rarely executed on a one-on-one apprenticeship basis. Instead, it is often conducted through a multiple-on-multiple basis, in which teams of faculty, postdoctoral fellows, graduate students, and even undergraduate students interact to create meaning from multiple but interconnected inquiry efforts (Charlesworth, Farrall, Stokes, & Turnbull, 1989; Delamont & Atkinson, 2001; Knorr-Cetina, 1999; Parry, 2007). In this context, the individual perspectives of faculty mentors may not be the best source of information regarding their mentees' research skills. It is possible that collective discussions among those who jointly contribute to a student's mentoring environment could provide a triangulated perspective that would more consistently align with performance-based measures. However, previous studies of distributed reasoning and recollection of individuals' contributions in research lab meetings suggest that this alternative may not yield reliable information either (Dunbar, 1997, 2000).

Student and faculty responses to questions about the nature of their interactions suggest that the aforementioned explanations are both valid and intertwined. For example, a first-year doctoral student in biology stated that guidance for research occurred within a series of relationships, which included, "[her] supervisor, another PhD student, a visiting professor, and undergraduates." This student initially described her relationship with her faculty advisor as "very hands-off. . . . He is a point of reference, but in my daily activities he is behind the scenes and not actively involved in what I am doing." Her faculty mentor validated this perception:

> We (the lab group) are pretty much trickle down. I have senior graduate students, senior undergraduate and so on. I am not directly involved in day-to-day activities. . . . I am pretty laissez faire when dealing with graduate students. My job is to bring money in, because what we do is expensive. So I view my role as to bring money in and give [the student] the opportunity to succeed . . . you can choose to take it or to not.

Another characteristic of the mentor-mentee relationship that may hinder faculty perceptions of students' skills is the manner in which work is expected to be done. As indicated previously, some mentors expect mentees to rely heavily on the guidance of other members of the research group. Even when this is not an explicit expectation, tasks are typically assigned to mentees with the expectation that they will work independently of the mentor's direct supervision and return when the work product is relatively complete and ready for review by the mentor. In this circumstance, the mentor does not have the opportunity to observe the process in which the mentee engages. For example, one participant commented that "the professor just throws you a topic or study, okay you're to do that and then you did it by yourself and then you give some information and you ask him [what more is needed]." To the extent that skills manifest as part of the research process without necessarily being directly assessable through the work product itself, mentors may draw unsubstantiated conclusions about the proficiency with which those aspects are performed.

Conversely, if students are working on predetermined projects (e.g., tasks necessary to the fulfillment of an awarded research grant), there may be a variety of skills that they do not have the opportunity to demonstrate due to project constraints. For example, the research design might be decided when a grant was written, and the mentee's role is to implement specified actions without applying his or her own relevant skills to the decision-making process. In one instance, a participant explained that his mentor "was the one who gave me the experimental design and then I wrote the experiments and then it was just discussion based." Similarly, another student explained that her mentor "had more control over the direction that it (the project) went, and more control over . . . specific cases, what cases we should look at."

Student Perceptions

These dynamics may similarly affect students' perceptions of their own abilities. Whereas faculty mentors may have more limited access to observe their mentees' skills than they or others assume, the mentees have fewer opportunities to receive targeted feedback to shape their self-assessments (Dunning, Heath, & Suls, 2004). Typically, as such opportunities accrue, alignment between students' self-assessments and the assessments of others improve to a limited extent (Falchikov & Boud, 1989). While other members of the laboratory community might provide them with assistance or guidance in completing tasks for which they are responsible, it is not guaranteed that such information constitutes direct feedback on the processes that the students do use or apply the standards that the faculty mentor would hold (e.g., Norcini, 2003). Studies of feedback practices in informal situations

and among peers indicate that messages are frequently softened to avoid conflict or emotional harm (Larson, 1989).

Another factor that may affect students' ability to self-assess is the lack of specifically articulated quality benchmarks for performance within the context of academic research. Various scholars studying the process of graduate education point to the implicit nature of both the standards for criterial tasks and overall performance expectations (e.g., Delamont & Atkinson, 2001). As Lovitts (2007) observes, "many of the standards are indeterminate qualities that faculty can recognize but not articulate precisely. . . . Indeed, it is not uncommon for faculty to respond to a protocol item by saying, 'you know it when you see it'" (p. 114). Empirical studies indicate that students frequently are better able to assess their own abilities and those of their peers when they make assessments of competence in relation to multiple, well-defined criteria for performance (Falchikov & Goldfinch, 2000; Norcini, 2003). However, one study of graduate students' abilities to assess the quality of research methodologies indicated that even when feedback was provided that detailed the flaws they had missed, their self-assessments of their ability to critique their own work successfully were unaffected (Caputo & Dunning, 2005)

Despite these challenges, one area of significant alignment between student interview- and performance-based assessments emerged: students' self-reported ability to generate literature reviews and their observed ability to set their research proposals in the context of other work in the field. Increased accuracy of self-assessment in this area is congruent with earlier work indicating that an awareness of the literature in one's field is one of the earliest skills to develop in STEM graduate students (Timmerman, Feldon, Maher, Strickland, & Gilmore, 2013). The generation of literature reviews is a task commonly given to less experienced students as a means to familiarize them with current development within a relevant field. For example, one mentor explained, "Basically, I assign him a topic and he is able to look at the literature to find the relevant background information." Another commented that "being able to look at the scientific literature and then develop—that's the first piece." For this reason, it is likely that this aspect of research skill is the one with which students have the most experience. It is also possible that it is a focal area for mentor-provided feedback, as faculty perceptions aligned with their mentees' perceptions in two of three instances in fall and one of one instance in spring. However, the poor alignment between faculty assessments and their mentees' performance-based assessments overall suggest limitations to that explanation.

Misaligned Perceptions

As noted earlier, most knowledge about the graduate mentoring process is typically based on data collected exclusively from doctoral students

without corresponding data from their mentors (Barnes & Austin, 2009). Thus, it is difficult to ascertain with any surety the extent to which perceptual misalignment found in this study was to be anticipated. However, studies of employee performance appraisals from the perspective of both the employee and direct supervisor might provide some cautionary insight. Low correlations between employee self-appraisals and supervisor appraisals (e.g., $r = .35$, Harris & Schaubroeck, 1988) have been consistently observed (Conway & Huffcutt, 1997; Thornton, 1968). Explanations abound and often include reference to contextual error sources (e.g., the "halo effect," Cooper, 1981; "similar-to-me effects," Wexley, Alexander, Greenawalt, & Couch, 1980). However, as Baruch (1996) has suggested, other commonly accepted explanations for low agreement between self and supervisor ratings include differences in the performance information available to different raters and variations in performance criteria used by different raters.

Within the pressing constraints and demands shaping the current doctoral education landscape, it is likely that mentors and mentees do in fact hold different perceptions of mentee skill level. The last common explanation, variations in rater performance criteria, also deserves consideration within the context of this study and speaks to both study limitations and future research directions. Study findings of misaligned perceptions were largely unanticipated within the context of the larger project examining the impacts of STEM graduate students' teaching and research experiences on the development of their research skills during the first years of their degree programs. A more fine-grained analysis of perceptual alignment might include, for example, providing interview questions to study participants in advance or using a combination of structured survey questions with interviews questions. Future studies in this area are encouraged to consider these suggestions.

Implications

Despite the findings, the intent of this analysis is not to discount the necessity of the mentor-mentee relationship. Many graduate students need faculty mentors to define the way forward to professional accomplishment (Boyle & Boice, 1998; Gardner, 2009; Kam, 1997). Likewise, faculty mentors, as stewards of their disciplines, need graduate student mentees to sustain and advance their fields (Golde & Walker, 2006). At a more mundane level, graduate students are also essential to accomplish the myriad of daily tasks that define the research process (Charlesworth et al., 1989; Knorr-Cetina, 1999; Pole, 2000).

However, study findings do call into question the extent to which, in today's academy, faculty mentors are able to fulfill their role within a cognitive apprenticeship model. The pressing need to "bring in money,"

combined with heightened expectations for scholarly productivity and frequent calls for teaching effectiveness, have profoundly affected the nature and pace of academic work (Peters & Olssen, 2005). Perhaps the key implication of this study is the challenge to deeply consider how these trends affect—and perhaps disrupt—the vital learning posited to occur within the mentor-mentee relationship at the graduate level.

As a more practical consideration, the failure to successfully triangulate mentor, mentee, and performance-based perspectives on students' research skills in this study presents a serious challenge to the common assumption that mentors' judgments about their mentees' abilities and training needs are an optimal source of guidance or assessment for students. Further, as a methodological consideration, the data reported here indicate that extreme caution should be used when relying on interview data that identify students' abilities or growth in terms of research competencies. Given that the ability to conduct independent research is a common benchmark of doctoral attainment, future research examining the quality of graduate training or its effectiveness in preparing new scholars should utilize independent evaluations of student skill that can be verified against performance-based sources of data (Feldon, Maher, et al., 2010).

With regard to educational practice, increasing the specificity of the learning objectives and the extent to which they are identified explicitly for STEM graduate students may enhance efficacy and precision of feedback offered by both faculty mentors and other members of laboratory research teams. Further, increased emphasis on anchoring assessments of skill to discrete task performance may enable faculty to more effectively steer students toward resources and experiences that will enhance necessary scholarly skills.

Notes

[1]Mean scores on some planks of the rubric did differ significantly by both institution and/or degree program, as reported in Feldon et al. (2011, p. S5). However, these differences were not reflected in the alignment outcomes of interest in the current study. Chi-square analyses were nonsignificant ($p > .05$). Given the limited sample sizes available for some analyses, visual inspection of the data was also utilized to search for noticeable trends that might not attain statistical significance. No such trends were identified.

References

Ahern, K., & Manathunga, C. (2004). Clutch-starting stalled research students. *Innovative Higher Education*, *28*, 237–254.

Åkerlind, G. S. (2005). Variation and commonality in phenomenographic research methods. *Higher Education Research and Development*, *24*, 321–334.

Anderson, M. S., & Swazey, J. P. (1998). Reflections on the graduate student experience: An overview. *New Directions for Higher Education*, *101*, 3–13.

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

Austin, A. E., & McDaniels, M. (2006). Preparing the professoriate of the future: Graduate student socialization for faculty roles. *Higher Education: Handbook of Theory and Research*, *21*, 397–456.

Barnes, B. J., & Austin, A. E. (2009). The role of doctoral advisors: A look at advising from the advisor's perspective. *Innovative Higher Education*, *33*, 297–315.

Baruch, Y. (1996). Self-performance appraisal vs direct-manager appraisal: A case of congruence. *Journal of Managerial Psychology*, *11*, 50–65.

Bishop, G. F., Tuchfarber, A., & Oldendick, R. W. (1986). Opinions on fictitious issues: The pressure to answer survey questions. *Public Opinion Quarterly*, *50*, 240–250.

Bloch, D. A., & Kraemer, H. C. (1989). 2 × 2 kappa coefficients: Measures of agreement or association. *Biometrics*, *45*, 269–287.

Boyle, P., & Boice, B. (1998). Best practices for enculturation: Collegiality, mentoring, and structure. *New Directions for Higher Education*, *101*, 87–94.

Bronfenbrenner, U. (1976). The experimental ecology of education. *Educational Researcher*, *5*, 5–15.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, *18*, 35–42.

Caputo, D., & Dunning, D. (2005). What you don't know: The role played by errors of omission in imperfect self-assessments. *Journal of Experimental Social Psychology*, *41*, 488–505.

Carnegie Initiative on the Doctorate. (2001). *Overview of doctoral educational studies and reports: 1990-present*. Stanford, CA: The Carnegie Foundation for the Advancement of Teaching.

Chan, T. F. (2006). A time for a change? The mathematics doctorate. In C. M. Golde, G. E. Walker, & Associate (Eds.), *Envisioning the future of doctoral education* (pp. 120–134). San Francisco, CA: Jossey-Bass.

Charlesworth, M., Farrall, L., Stokes, T., & Turnbull, D. (1989). *Life among the scientists*. Melbourne: Oxford University Press.

Cicchetti, D. V. (1991). The reliability of peer review for manuscript and grant submissions: A cross-disciplinary investigation. *Behavioral and Brain Sciences*, *14*, 119–135.

Clark, R., Harden, S., & Johnson, W. (2000). Mentor relationships in clinical psychology doctoral training: Results of a national survey. *Teaching of Psychology*, *27*, 262–268.

Collins, A., Brown, J. S., & Holum, A. (1991). Cognitive apprenticeship: Making thinking visible. *American Educator*, *6*, 38–46.

Conway, J. M., & Huffcutt, A. I. (1997). Psychometric properties of multisource performance ratings: A meta-analysis of subordinate, supervisor, peer, and self-ratings. *Human Performance*, *10*, 331–360.

Cooper, W. H. (1981). Ubiquitous halo. *Psychological Bulletin*, *90*, 218–244.

Cox, M., & Andriot, A. (2009). Mentor and undergraduate student comparisons of students' research skills. *Journal of STEM Education*, *10*, 31–39.

Creswell, J. W., Plano Clark, V. L., Gutmann, M. L., & Hanson, W. E. (2003). Advanced mixed methods research designs. In A. Tashakkori & C. Teddlie (Eds.),

Handbook of mixed methods in social & behavioral research (pp. 209–240). Thousand Oaks, CA: Sage Publications.

Cumming, J. (2009). The doctoral experience in science: Challenging the current orthodoxy. *British Educational Research Journal*, *35*, 877–890.

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

Delamont, S., Atkinson, P., & Parry, O. (2000). *The doctoral experience: Success and failure in graduate school.* London: Falmer Press.

Denicolo, P. (2003). Assessing the PhD: A constructive view of criteria. *Quality Assurance in Education*, *11*, 84–91.

Dunbar, K. (1997). How scientists think: On-line creativity and conceptual change in science. In T. B. Ward & S. M. Smith (Eds.), *Creative thought: An investigation of conceptual structures and processes* (pp. 461–493). Washington, DC: American Psychological Association.

Dunbar, K. (2000). How scientists think in the real world: Implications for science education. *Journal of Applied Developmental Psychology*, *21*, 49–58.

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest*, *5*, 69–106.

Dunning, D., Johnson, K., Ehrlinger, J., & Kriger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science*, *12*, 83–87.

Ehrlinger, J., & Dunning, D. (2003). How chronic self-views influence (and potentially mislead) estimates of performance. *Journal of Personality and Social Psychology*, *84*, 5–17.

Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research*, *59*, 395–430.

Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research*, *70*, 287–322.

Feldon, D. F. (2007). Cognitive load and classroom teaching: The double-edged sword of automaticity. *Educational Psychologist*, *42*, 123–137.

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.

Feldon, D. F., Maher, M., & Timmerman, B. (2010). A call for performance-based data in the study of STEM graduate education. *Science*, *329*, 282–283.

Feldon, D. F., Peugh, J., Timmerman, B. E., Maher, M. A., Hurst, M., Strickland, D., Gilmore, J. A., & Stiegelmeyer, C. (2011). Graduate students' teaching experiences improve their methodological research skills. *Science*, *333*, 1037–1039.

Feldon, D. F., Timmerman, B. E., Stowe, K., & Showman, R. (2010). Translating expertise into effective instruction: The impacts of cognitive task analysis (CTA) on lab report quality and student retention in the biological sciences. *Journal of Research in Science Teaching*, *47*, 1165–1185.

Fleiss, J. L., & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, *33*, 613–619.

Franke, A., & Arvidsson, B. (2011). Research supervisors' different ways of experiencing supervision of doctoral students. *Studies in Higher Education*, *36*, 7–19.

Gardner, S. K. (2007). "I heard it through the grapevine": Doctoral student socialization in chemistry and history. *Higher Education*, *54*, 723–740.

Gardner, S. K. (2008). "What's too much and what's too little?": The process of becoming an independent researcher in doctoral education. *Journal of Higher Education*, *79*, 326–350.

Gardner, S. K. (2009). Conceptualizing success in doctoral education: Perspectives of faculty in seven disciplines. *Review of Higher Education*, *32*, 383–406.

Gilmore, J., Strickland, D., Timmerman, B., Maher, M., & Feldon, D. F. (2010). Weeds in the flower garden: An exploration of plagiarism in graduate students' research proposals and its connection to enculturation, ESL, and contextual factors. *International Journal of Educational Integrity*, *6*, 13–28.

Glaser, B. G. (1965). The constant comparative method of qualitative analysis. *Social Problems*, *12*, 436–445.

Golde, C. M. (1998). Beginning graduate school: Explaining first-year doctoral attrition. *New Directions for Higher Education*, *101*, 55–64.

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. (2001). *At cross purposes: What the experiences of doctoral students reveal about doctoral education*. Philadelphia, PA: The Pew Charitable Trusts.

Golde, C. M., & Walker, G. E. (Eds.). (2006). *Envisioning the future of doctoral education: Preparing stewards of the discipline*. San Francisco, CA: Jossey-Bass.

Goldstein, S. D., Lindeman, B., Colbert-Getz, J., Arbella, T., Dudas, R., Lidor, A., & Sacks, B. (2014). Faculty and resident evaluations of medical students on a surgery clerkship correlate poorly with standardized exam scores. *The American Journal of Surgery*, *207*, 231–235.

Green, S. G., & Bauer, T. N. (1995). Supervisory mentoring by advisers: Relationships with doctoral student potential, productivity, and commitment. *Personnel Psychology*, *48*, 537–562.

Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*, *11*, 255–274.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*, 4–27.

Haaga, D. A. F. (1993). Peer review of term papers in graduate psychology courses. *Teaching of Psychology*, *20*, 28–32.

Halonen, J. S., Bosack, T., Clay, S., McCarthy, M., Dunn, D. S., Hill, G. W. IV, ... Whitlock, K. IV. (2003). A rubric for learning, teaching, and assessing scientific inquiry in psychology. *Teaching of Psychology*, *30*, 196–208.

Harris, M. M., & Schaubroeck, J. (1988). A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. *Personnel Psychology*, *41*, 43–62.

Hawley, P. (1993). *Being bright is not enough: The unwritten rules of doctoral studies*. Springfield, IL: Charles C. Thomas.

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.

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.

Johnson, L., Lee, A., & Green, B. (2000). The Ph.D. and the autonomous self: Gender, rationality, and postgraduate pedagogy. *Studies in Higher Education*, *25*, 135–147.

Johnson, R. L., Penny, J., & Gordon, B. (2009). *Assessing performance: Designing, scoring, and validating performance tasks*. New York, NY: Guilford Press.

Johnson, R. L., Penny, J., Gordon, B., Shumate, S. R., & Fisher, S. P. (2005). Resolving score differences in the rating of writing samples: Does discussion improve the accuracy of scores? *Language Assessment Quarterly: An International Journal*, *2*, 117–146.

Johnson, W. (2002). The intentional mentor: strategies and guidelines for the practice of mentoring. *Professional Psychology: Research and Practice*, *33*, 88–96.

Kam, B. H. (1997). Style and quality in research supervision: The supervisor dependency factor. *Higher Education*, *34*, 81–103.

Kardash, C. (2000). Evaluation of an undergraduate research experiences: perceptions of undergraduate interns and their faculty mentors. *Journal of Educational Psychology*, *92*, 191–201.

Kelly, G. J., & Takao, A. (2002). Epistemic levels in argument: An analysis of university oceanography students' use of evidence in writing. *Science Education*, *86*, 314–342.

Knorr-Cetina, K. (1999). *Epistemic cultures: How the sciences make knowledge*. Cambridge, MA: Harvard University Press.

Krosnick, J. A., Holbrook, A. L., Berent, M. K., Carson, R. T., Hanneman, W. M., Kopp, R. J, . . . Conaway, M. (2002). The impact of "no opinion" response options on data quality: Non-attitude reduction or an invitation to satisfice? *Public Opinion Quarterly*, *66*, 371–403.

Lagowski, J., & Vick, J. (1995). Faculty as mentors. *New Direction for Teaching and Learning*, *62*, 79–85.

Larson, J. R., Jr. (1989). The dynamic interplay between employees' feedback-seeking strategies and supervisors' delivery of performance feedback. *Academy of Management Review*, *14*, 408–422.

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

Lovitts, B. E. (2007). *Making the implicit explicit: Creating performance expectations for the dissertation*. Sterling, VA: Stylus Publishing.

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

Macrae, C. N., Bodenhausen, G. V., Milne, A. B., & Wheeler, V. (1996). On resisting the temptation for simplification: Counterintentional effects of stereotype suppression on social memory. *Social Cognition*, *14*, 1–20.

Maher, M., Gilmore, J. A., Feldon, D. F., & Davis, T. (2014, April). *Doctoral student mentoring and the role of cognitive apprenticeship*. Paper presented at the 2014 meeting of the American Educational Research Association, Philadelphia, PA.

Marsh, H. W., & Ball, S. (1989). The peer review process used to evaluate manuscripts submitted to academic journals: Interjudgmental reliability. *Journal of Experimental Education*, *57*, 151–169.

Marsh, H. W., & Bazeley, P. (1999). Multiple evaluations of grant proposals by independent assessors: Confirmatory factor analysis evaluations of reliability, validity, and structure. *Multivariate Behavioral Research*, *34*, 1–30.

Marton, F., & Pong, W. Y. (2005). On the unit of description in phenomenography. *Higher Education Research & Development*, *24*, 335–348.

Merton, R. K. (1973). *The sociology of science*. Chicago, IL: University of Chicago Press.

Milem, J. F., Berger, J. B., & Dey, E. L. (2000). Faculty time allocation: A study of change over twenty years. *Journal of Higher Education*, *71*, 545–475.

Nettles, M. T., & Millett, C. M. (2006). *Three magic letters: Getting to Ph.D.* Baltimore, MD: John Hopkins University Press.

Nickerson, R. S. (1999). How we know—and sometimes misjudge—what others know: Imputing one's own knowledge to others. *Psychological Bulletin*, *125*, 737-159.

Norcini, J. (2003). Peer assessment of competence. *Medical Education*, *37*, 539–543.

Paglis, L., Green, S., & Bauer, T. (2006). Does mentor mentoring add value? A longitudinal study of mentoring and doctoral student outcomes. *Research in Higher Education*, *47*, 451–476.

Parry, S. (2007). *Disciplines and the doctorate*. Dordrecht, the Netherlands: Springer.

Pearson, M., & Brew, A. (2002). Research training and supervision development. *Studies in Higher Education*, *27*, 135–150.

Peters, M. A., & Olssen, M. (2005). Useful knowledge: Redefining research and teaching in the learning economy. In R. Barnett (Ed.), *Reshaping the university: New relationships between research, scholarship and teaching* (pp. 37–47). Buckingham: Society for Research into Higher Education/Open University Press.

Petty, R. E., Fleming, M. A., & Fabrigar, L. R. (1999). The review process at PSPB: Correlates of interreviewer agreement and manuscript acceptance. *Personality and Social Psychology Bulletin*, *25*, 188–203.

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

Pole, C. J., Sprokkereef, A., Burgess, R. G., & Lakin, E. (1997). Supervision of doctoral students in the natural sciences: Expectations and experiences. *Assessment and Evaluation in Higher Education*, *22*, 49–63.

Remler, D. K., & Pema, E. (2009). *Why do institutions of higher education reward research while selling education?* (NBER Working Paper No. 14974). Cambridge, MA: National Bureau of Economic Research.

Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist*, *54*, 95–105.

Sloman, S. A. (2002). Two systems of reasoning. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 379–396). New York, NY: Cambridge University Press.

Stajkovic, A. D., & Luthans, F. (1998). Self-efficacy and work-related performance: A meta-analysis. *Psychological Bulletin*, *124*, 240–261.

Svensson, L. (1997). Theoretical foundations of phenomenography. *Higher Education Research & Development*, *16*, 159–171.

Tariq, V. N., Stefan, L. A. J., Butcher, A. C., & Heylings, D. J. A. (1998). Developing a new approach to the assessment of project work. *Assessment & Evaluation in Higher Education*, *23*, 221–240.

Thornton, G. C. (1968). The relationship between supervisory- and self-appraisals of executive performance. *Personnel Psychology*, *21*, 441–455.

Timmerman, B. E. (2008). *Peer review in an undergraduate biology curriculum: Effects on students' scientific reasoning, writing, and attitudes* (Unpublished doctoral dissertation). Curtin University of Technology, Perth, Australia.

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.

Timmerman, B. E., Johnson, R. L., & Payne, J. (2007, April). *Development of a universal rubric for assessing students' science inquiry skills.* National Association of Research in Science Teaching 2007 Annual Meeting, New Orleans LA.

Timmerman, B. E. C., Strickland, D. C., Johnson, R. L., & Payne, J. R. (2011). Development of a "universal" rubric for assessing undergraduates' scientific

reasoning skills using scientific writing. *Assessment & Evaluation in Higher Education*, *36*, 509–547.

Topping, K. J., Smith, E. F., Swanson, I., & Elliot, A. (2000). Formative peer assessment of academic writing between postgraduate students. *Assessment & Evaluation in Higher Education*, *25*, 149–169.

Waldeck, J. H., Orrego, V. O., Plax, T. G., & Kearney, P. (1997). Graduate student/faculty mentoring relationships: Who gets mentored, how it happens, and to what end. *Communication Quarterly*, *45*, 93–109.

Walker, G. E., Golde, C., Jones, L., Conklin Bueschel, A., & Hutchings, P. (2008). *The formation of scholars: Rethinking doctoral education for the twenty-first century.* San Francisco, CA: Jossey-Bass.

Weidman, J. C., Twale, D. J., & Stein, E. L. (2001). *Socialization of graduate and professional students: A perilous passage?* (ASHE-ERIC Higher Education Report 28, no. 3). Washington, DC: Association of the Study of Higher Education.

Wexley, K. N., Alexander, R. A., Greenawalt, J. P., & Couch, M. A. (1980). Attitudinal congruence and similarity as related to interpersonal evaluations in manager-subordinate dyads. *Academy of Management Journal*, *23*, 320–330.

Willison, J., & O'Regan, K. (2007). Commonly known, commonly not known, totally unframed: A framework for students becoming researchers. *Higher Education Research & Development*, *26*, 393–409.

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.

Zhao, C., Golde, C. M., & McCormick, A. C. (2007). More than a signature: How advisor choice and advisor behavior affect doctoral student satisfaction. *Journal of Further and Higher Education*, *31*, 263–281.

Zimmerman, C. (2000). The development of scientific reasoning. *Developmental Review*, *20*, 99–149.