Psychological Research (2021) 85:1108–1113
https://doi.org/10.1007/s00426-020-01307-2

ORIGINAL ARTICLE

Toward a cumulative science of expertise: commentary on Moxley, Ericsson, and Tuffiash (2017)

Brooke N. Macnamara[1] · **David Z. Hambrick**[2]

Received: 28 November 2018 / Accepted: 13 February 2020 / Published online: 9 March 2020

Abstract
In a recent *Psychological Research* article, Moxley, Ericsson, and Tuffiash (2017) report two studies of SCRABBLE expertise. The results revealed that the average SCRABBLE rating was higher for males than for females. Moreover, correlational and structural equation analyses revealed that activities that the authors refer to as "purposeful practice" accounted for a substantial amount of the variance in SCRABBLE ratings. The authors generalize their findings concerning SCRABBLE to STEM careers. We believe this generalization is unjustified, as is their argument that SCRABBLE can be used to understand the gender gap in STEM fields. Moreover, the authors' conclusions are undermined by inconsistencies and contradictions in their arguments. We discuss these problems with Moxley et al.'s article in terms of their impact on the cumulative science of expertise.

Introduction

There is currently a great deal of interest in the origins of individual differences in domain-specific performance (i.e., expertise). In a recent *Psychological Research* article, Moxley, Ericsson, and Tuffiash (2017) report two studies of SCRABBLE expertise. In each study, participants recruited from the National SCRABBLE Championship completed a survey with questions about their engagement in SCRABBLE-relevant practice activities. The results revealed that the average SCRABBLE rating was higher for males than for females. Moreover, correlational and structural equation analyses revealed that activities that the authors refer to as "purposeful practice" accounted for a substantial amount of the variance in SCRABBLE ratings.

We credit the authors for this work. Their research has notable strengths. The studies used relatively large samples of performers representing wide and continuous ranges of expertise, as assessed by an objective measure of skill (i.e., SCRABBLE rating). At the same time, we believe that the authors' generalizations of their findings concerning SCRABBLE to STEM careers are unjustified, as is their argument that SCRABBLE can be used to understand the gender gap in STEM fields. Moreover, the authors' conclusions are undermined by inconsistencies and contradictions in their arguments. We are primarily concerned with the authors' use of what can only be called *theoretical term swapping* and the negative impact this has on the cumulative science of expertise. We briefly discuss these issues in turn.

Using SCRABBLE to understand the STEM gender gap

The authors open their article with the observation that "[g]ender gaps in academia, particularly in STEM fields, are currently of great national concern" (Moxley et al., 2017, p. 1). They are indeed, and attempting to shed light on this important issue, the authors suggest that SCRABBLE is a performance domain "requiring reasonably similar abilities to those necessary for success in professional STEM domains" (Moxley et al., 2017, p. 2). But how are the abilities needed to play SCRABBLE—a word game—"reasonably similar" to the abilities needed to be a successful mathematician or engineer? The authors do not say. If anything, the abilities needed for success in SCRABBLE probably have little overlap with the abilities needed for success in STEM fields. For example, the only math needed in SCRABBLE is simple addition to track scores.

✉ Brooke N. Macnamara
brooke.macnamara@case.edu

[1] Department of Psychological Sciences, Case Western Reserve University, Cleveland, USA

[2] Michigan State University, East Lansing, USA

Nevertheless, generalizing their findings from SCRABBLE to the STEM gender gap, the authors report that "ratings of enjoyment explain the gender difference in the behavior [purposeful practice] which best predicted SCRABBLE skill" (Moxley et al., 2017, p. 15) and contend that their "findings are relevant to the study of individual differences in a large number of recreational and professional activities" (Moxley et al., 2017, p. 16). More specifically, based on their SCRABBLE findings, the authors argue that women find less success relative to men in STEM because women are less inclined than men to engage in the most beneficial—and least enjoyable—types of practice activities. In their own words, "we have proposed how large gender differences can be attributed to differences in the methods of skill acquisition…. We found that those differences appear to be due to preferences for engaging in certain types of domain-related activities" (Moxley et al., 2017, p. 19).

However, motivational influences on how people approach SCRABBLE versus a career are surely different. Perhaps women engage in less unenjoyable SCRABBLE practice than men because they think to do so is silly and see SCRABBLE for what it is: a board game designed to be fun. Or maybe the women in the authors' samples had more stressful careers than the men, and after long, hard days at the office simply wanted to have fun playing competitive SCRABBLE. These speculations seem at least as well founded as generalizing findings from SCRABBLE to STEM.

Continuing with the notion that the mean gender difference in competitive SCRABBLE is a reasonable corollary to the gender gap in STEM fields, the authors state they believe domains such as SCRABBLE provide opportunities to study "how preferences, personality and, in our opinion most importantly, behavioral differences in the engagement in effective practice activities can produce a gender gap" (Moxley et al., 2017, p. 19). Engagement in effective practice activities could, in principle, produce a gender gap. However, a major source of the gender gap in STEM is that girls do not pursue STEM careers in the first place (e.g., Sadler, Sonnert, Hazari, & Tai, 2012; Sax & Arms, 2008); the number one-cited reason for this is a lack of role models (see e.g. Bauer & Microsoft, 2017, and Kesar, Microsoft, & KRC Research, 2018, for results of interviews with over 17,000 girls and women). The authors also fail to mention the extremely high rates of sexual harassment in STEM fields that lead to women leaving these jobs (e.g., National Academies of Sciences, Engineering, and Medicine, 2018). At a more general level, the authors seem not to appreciate the multifactorial nature of the STEM gender gap (see, e.g., De Welde & Laursen, 2011).

Shifting theoretical definitions

As a theoretical framework for their studies, the authors distinguish between two forms of practice (see also Ericsson & Pool, 2016). They describe *purposeful practice* as self-directed training "in which the individual engages in practice tasks with opportunities for feedback, repetition, and refinement" (Moxley et al., 2017, p. 3), whereas *deliberate practice* has the additional requirement "that the individual meets regularly with a teacher, who can assess his or her current performance level, recommend appropriate targets for improvement and then describe training activities that the individual can engage in by themselves with opportunities for feedback, repetition, and refinement" (Moxley et al., 2017, p. 3). As Ericsson and Pool (2016) noted when they introduced this distinction, "we are drawing a clear distinction between purposeful practice—in which a person tries very hard to push himself or herself to improve—and practice that is both purposeful and *informed*" (p. 98).

We have no issue with the new distinction between purposeful practice and deliberate practice. Research aimed at better understanding the "microstructure" (Ericsson, 1996) of domain-specific experience may yield important discoveries about sources of individual differences in expertise. The problem lies in how the authors use the terms, in view of how they have used them in past writings (in co-authored articles and articles with others). First and foremost, the authors have defined deliberate practice in flatly contradictory ways. In particular, while in the present article (and in some other articles), the authors explain that a teacher or coach must design a training activity for it to qualify as deliberate practice, in the past they have been inconsistent on this critical point (see Table 1, Fig. 1). For example, Ericsson (1998) explained in clear terms that deliberate practice activities do *not* need to be designed by a teacher, stating that "Ericsson et al. (1993) proposed the term deliberate practice to refer to those training activities that were designed solely for the purpose of improving individuals' performance by a teacher *or the performers themselves*" (p. 84, emphasis added).

As an especially striking—and perplexing—illustration of this point, Moxley et al. (2017) explain that "SCRABBLE players cannot engage in deliberate practice, but only purposeful practice and other types of practice" (p. 4; see also Table 2 below). However, in a previous article, two of the authors (Tuffiash and Ericsson) conducted a study of SCRABBLE and concluded that expertise "was significantly associated with the quantity of time spent on SCRABBLE-related activities that *best met the theoretical description of deliberate practice*" (Tuffiash, Roring, & Ericsson, 2007, p. 131, emphasis added). And, citing that

Table 1 The shifting definition of deliberate practice

Year	Definition	Quote	References
1996	Requires a teacher	"Ericsson et al. (1993) defined deliberate practice as a very specific activity designed for an individual by a skilled teacher explicitly to improve performance."	Krampe and Ericsson (1996, p. 333)
1998	Does not require a teacher	"Ericsson et al. (1993) proposed the term deliberate practice to refer to those training activities that were designed solely for the purpose of improving individuals' performance by a teacher or the performers themselves."	Ericsson (1998, p. 84)
2000	Requires a teacher	"Ericsson et al. (1993) identified activities that met the necessary requirements for effective training and were designed by a teacher to improve a specific individual's performance. They termed these activities 'deliberate practice.'"	Ericsson (2000, p. 368)
2007	Does *not* require a teacher	"Ericsson et al. (1993) introduced the term deliberate practice to describe focused and effortful practice activities that are pursued with the explicit goal of performance improvement. Deliberate practice implies that well-defined tasks are practiced at an appropriate level of difficulty and that informative feedback is given to monitor improvement. These activities can be designed by external agents, such as teachers or trainers, or by the performers themselves."	Keith and Ericsson (2007, p. 136)
2015	Requires a teacher	"When this type of training is supervised and guided by a teacher, it is called' deliberate practice'—a concept my colleagues and I introduced in 1993."	Ericsson (2015, p. 1472)

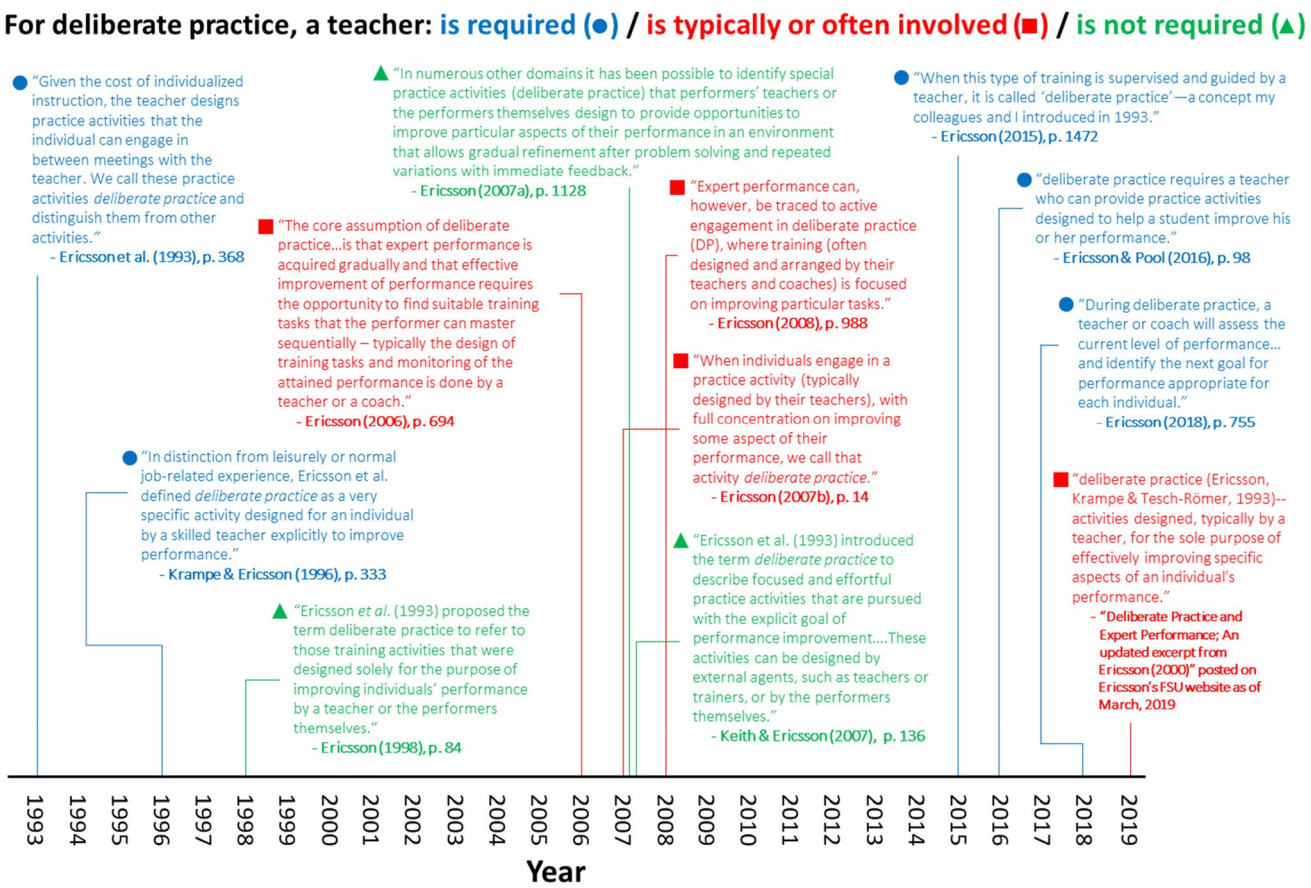

Fig. 1 Chronology of definitions of deliberate practice. Definitions in blue (filled circle) indicate a teacher is required to design deliberate practice activities. Definitions in red (filled square) indicate a teacher is typically or often involved in designing deliberate practice activities. Definitions in green (filled triangle) indicate a teacher is not required to design deliberate practice activities (color figure online)

Table 2 Instances of term swapping and shifting in Moxley et al.'s (2017) article

Study	Claim by Moxley et al. (2017) concerning "purposeful practice"—a concept distinct from deliberate practice	Earlier claim that the same study concerned "deliberate practice"
Charness et al. (2005)	"Charness et al. (2005) found evidence for an independent effect of engagement in *purposeful practice* for chess skill, even after controlling for other types of practice activities" (p. 17)	"This paper (Charness et al., 2005) reports the most compelling and detailed evidence for how designed training (*deliberate practice*) is the crucial factor in developing expert chess performance" (Ericsson, 2005, p. 237)
Duckworth et al. (2011)	"…measures that Duckworth et al. (2011) had found to be related to *purposeful practice* in preparation for competitions in spelling" (p. 12)	Duckworth et al. (2011) "includes *deliberate practice* in its title, '*Deliberate practice* spells success.' In that study, we (as I was also one of the co-authors) collected data on '*deliberate practice*.' (Ericsson, 2012, p. 6)
Tuffiash, Roring, and Ericsson (2007)	"SCRABBLE players cannot engage in deliberate practice, but only *purposeful practice* and other types of practice… Tuffiash, Roring, and Ericsson (2007) assessed separate estimates of the amount of studying and also playing SCRABBLE" (p. 4)	"SCRABBLE expertise…was significantly associated with the quantity of time spent on SCRABBLE-related activities that best met the theoretical description of *deliberate practice*" (Tuffiash, Roring, & Ericsson, 2007, p. 131)
Tuffiash, Roring, and Ericsson (2007)	"SCRABBLE players cannot engage in deliberate practice, but only *purposeful practice* and other types of practice…. Tuffiash, Roring, and Ericsson (2007) assessed separate estimates of the amount of studying and also playing SCRABBLE" (p. 4)	"Several researchers have reported a consistent association between the amount and quality of solitary activities meeting the criteria of *deliberate practice* and performance in different domains of expertise, such as… Scrabble (Tuffiash et al., 2007)" (Ericsson et al., 2009, p. 9)

Instances of "purposeful practice" in quotations in the center column and "deliberate practice" in quotations in the right-hand column are italicized for emphasis.

study, Ericsson et al. (2009) explained that "researchers have reported a consistent association between the amount and quality of solitary activities *meeting the criteria of deliberate practice* and performance in different domains of expertise, such as…Scrabble (Tuffiash et al., 2007)" (p. 9, emphasis added). Apparently, the criteria for deliberate practice have shifted.

It is also perplexing that Moxley et al. (2017) claim that they collected data on "purposeful practice," given that this theoretical term had not yet been introduced at the time the data were collected. That is, as Moxley et al. note, their data were collected at the National SCRABBLE Championship in 2004 (Study 1) and 2008 (Study 2); however, Ericsson and Pool (2016) did not introduce the term "purposeful practice" until over a decade later. Moxley et al.'s claim that they measured purposeful practice is especially confusing given that the Study 1 data were collected concurrently with data reported in the Tuffiash et al. (2007) article focusing on "deliberate practice."

Theoretical term swapping

Without acknowledgement, the authors also replace the term 'deliberate practice' with 'purposeful practice' in describing some of their own and others' past research on deliberate practice (see Table 2). As a case in point, referring to a study of Spelling Bee contestants by Ericsson and colleagues (Duckworth, Kirby, Tsukayama, Berstein, & Ericsson, 2011), they explain that participants were asked to "fill out several additional personality measures that Duckworth et al. (2011) had found to be related to *purposeful practice* in preparation for competitions in spelling" (p. 12, emphasis added). However, that study concerned *deliberate practice*, not *purposeful practice*. This is clear from reading Duckworth et al. (2011) because (a) there is no mention of purposeful practice in the article, whereas the term "deliberate practice" appears in the title of the article, as a keyword, and 72 other times in the article; (b) the practice factor is labeled "deliberate practice" (to which one of the two personality measures Duckworth et al. reported was found to be related), and; (c) the major conclusion of the study was that "[d]eliberate practice time predicted performance in final competition" (Duckworth et al., 2011, p. 176). Furthermore, in emphatic terms, Ericsson (2012) previously argued that the Duckworth et al. study collected data on "deliberate practice," explaining that the report of the study "includes deliberate practice in its title, 'Deliberate practice spells success.' In that study, we (as I was also one of the co-authors) collected data on 'deliberate practice'" (p. 6).

As another example of this term swapping, referring to a study of chess by Charness, Tuffiash, Krampe, Reingold, and Vasyukova (2005), the authors state that "Charness et al.

(2005) found evidence for an independent effect of engagement in *purposeful practice* for chess skill, even after controlling for other types of practice activities" (p. 17, emphasis added). However, the term "purposeful practice" does not appear in the Charness et al. article; the title of the paper is "The Role of Deliberate Practice in Chess Expertise." Furthermore, Ericsson and Moxley (2012) previously explained that Charness et al. used an "elaborate interview procedure" to measure "number of hours of deliberate practice" (p. 652), and Ericsson (2005) concluded that the Charness et al. study "reports the most compelling and detailed evidence for how designed training (*deliberate practice*) is the crucial factor in developing expert chess performance" (p. 237, emphasis added). The authors do not acknowledge, much less justify, these shifts in theoretical terms.

The cumulative science of expertise

Why is it wrong for scientists to shift theoretical definitions and criteria, and to swap a term having one definition for a term having another definition, without explaining to their readers why they did so? It is wrong because it impedes theoretical progress in an area of scientific research, hindering the accumulation of knowledge toward a greater understanding of some phenomenon by empirically testing theories. That is, when a concept is defined in conflicting ways, the theory can accommodate any finding, making it unfalsifiable. A theorist can accept a finding under one definition or reject the same finding under another definition depending on whether it supports their theory; it is no longer possible to test the theory against a competing theory, and thus to advance knowledge. When a term having one definition is swapped for a term having another definition, a theorist can effectively rewrite history, erasing findings that pose a problem for their theory. What could we ever learn through scientific research if, as a universal practice, scientists felt free to shift their theoretical definitions and criteria without explanation, and to swap terms having different definitions?

In the present case, Moxley et al.'s (2017) findings are directly relevant to the ongoing debate about the importance of deliberate practice. The major finding from our own and others' research on this topic is that deliberate practice, while certainly important, leaves a large amount of the inter-individual variability in expertise unexplained and potentially explainable by other factors (e.g., Macnamara, Hambrick, & Oswald, 2014). To put it another way, deliberate practice is an important piece of the expertise puzzle, just not the only important piece. The authors' findings add to the case. However, they now call their measures of deliberate practice "purposeful practice," thwarting attempts by other researchers to empirically test the importance of deliberate practice.

We make no judgement about the motives behind the preceding actions—that is, whether or not they were intentional. We note simply that these actions are antithetical to a cumulative science of expertise.

Compliance with ethical standards

Conflict of interest Brooke N. Macnamara declares that she has no conflict of interest. David Z. Hambrick declares that he has no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Bauer, M. W., & Microsoft, Inc. (2017). *Why Europe's girls aren't studying STEM.* (White paper). Microsoft Philanthropies Europe.

Charness, N., Tuffiash, M., Krampe, R., Reingold, E., & Vasyukova, E. (2005). The role of deliberate practice in chess expertise. *Applied Cognitive Psychology, 19*, 151–165. https://doi.org/10.1002/acp.1106.

De Welde, K., & Laursen, S. (2011). The glass obstacle course: Informal and formal barriers for women Ph.D. students in STEM fields. *International Journal of Gender, Science and Technology, 3*, 571–595.

Duckworth, A. L., Kirby, T. A., Tsukayama, E., Berstein, H., & Ericsson, K. A. (2011). Deliberate practice spells success: Why grittier competitors triumph at the National Spelling Bee. *Social Psychological and Personality Science, 2*, 174–181. https://doi.org/10.1177/1948550610385872.

Ericsson, K. A. (1996). The acquisition of expert performance: An introduction to some of the issues. In K. A. Ericsson (Ed.), *The road to excellence: The acquisition of expert performance in the arts and sciences, sports, and games* (pp. 1–50). New York: Lawrence Erlbaum Associates.

Ericsson, K. A. (1998). The scientific study of expert levels of performance: General implications for optimal learning and creativity. *High Ability Studies, 9*, 75–100. https://doi.org/10.1080/1359813980090106.

Ericsson, K. A. (2000). How experts attain and maintain superior performance: Implications for the enhancement of skilled performance in older individuals. *Journal of Aging and Physical Activity, 8*, 366–372.

Ericsson, K. A. (2005). Recent advances in expertise research: A commentary on the contributions to the special issue. *Applied Cognitive Psychology, 19*, 233–241. https://doi.org/10.1002/acp.1111.

Ericsson, K. A. (2012). *The danger of delegating education to journalists: Why the APS Observer needs peer review when summarizing new scientific developments.* Unpublished manuscript. https://psy.fsu.edu/faculty/ericssonk/ericsson.hp.html. Accessed 5 Mar 2020.

Ericsson, K. A. (2015). Acquisition and maintenance of medical expertise: A perspective from the expert-performance approach with deliberate practice. *Academic Medicine, 90*, 1471–1486. https://doi.org/10.1097/ACM.0000000000000939.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*, 363–406. https://doi.org/10.1037/0033-295X.100.3.363.

Ericsson, K. A., & Moxley, J. A. (2012). A critique of Howard's argument for innate limits in chess performance or why we need an account based on acquired skill and deliberate practice. *Applied Cognitive Psychology, 26*, 649–653. https://doi.org/10.1002/acp.2841.

Ericsson, K. A., Perez, R. S., Eccles, D. W., Lang, L., Baker, E. L., Bransford, J. D., et al. (2009). The measurement and development of professional performance: An introduction to the topic and a background to the design and origin of this book. In K. A. Ericsson (Ed.), *Development of professional expertise: Toward measurement of expert performance and design of optimal learning environments* (pp. 1–24). Cambridge: Cambridge University Press.

Ericsson, A., & Pool, R. (2016). *Peak: Secrets from the new science of expertise*. New York: Houghton Mifflin Harcourt Publishing.

Keith, N., & Ericsson, K. A. (2007). A deliberate practice account of typing proficiency in everyday typists. *Journal of Experimental Psychology: Applied, 13*, 135–145. https://doi.org/10.1037/1076-898X.13.3.135.

Kesar, S., & Microsoft, Inc., & KRC Research. (2018). *Closing the STEM Gap: Why STEM classes and careers still lack girls and what we can do about it (White paper)*. Washington, DC: Microsoft Philanthropies.

Krampe, R. T., & Ericsson, K. A. (1996). Maintaining excellence: Deliberate practice and elite performance in young and older pianists. *Journal of Experimental Psychology: General, 125*, 331–359. https://doi.org/10.1037/0096-3445.125.4.331.

Macnamara, B. N., Hambrick, D. Z., & Oswald, F. L. (2014). Deliberate practice and performance in music, games, sports, education, and professions: A meta-analysis. *Psychological Science, 25*, 1608–1618. https://doi.org/10.1177/0956797614535810.

Moxley, J. A., Ericsson, K. A., & Tuffiash, M. (2017). Gender differences in SCRABBLE performance and associated engagement in purposeful practice activities. *Psychological Research Psychologische Forschung*. https://doi.org/10.1007/s00426-017-0905-3.

Sadler, P. M., Sonnert, G., Hazari, Z., & Tai, R. (2012). Stability and volatility of STEM career interest in high school: A gender study. *Science Education, 96*, 411–427. https://doi.org/10.1002/sce.21007.

Sax, L. J., & Arms, E. (2008). Gender differences over the span of college: Challenges to achieving equity. *NASPA Journal About Women in Higher Education, 1*, 25–50. https://doi.org/10.2202/1940-7890.1003.

The National Academy of Sciences, Engineering, and Medicine. (2018). *Sexual harassment of women: Climate, culture, and consequences in academic sciences, engineering, and medicine*. (Consensus study report). Washington, DC: The National Academies Press. https://doi.org/10.17226/24994

Tuffiash, M., Roring, R. W., & Ericsson, K. A. (2007). Expert performance in SCRABBLE: Implications for the study of the structure and acquisition of complex skills. *Journal of Experimental Psychology: Applied, 13*, 124–134. https://doi.org/10.1037/1076-898X.13.3.124.