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,* 0956797614535810. doi:10.1177/0956797614535810

Deliberate Practice and Performance in Music, Games, Sports, Education, and Professions: A Meta-Analysis

Brooke N. Macnamara[1]
David Z. Hambrick[2]
Frederick L. Oswald[3]

[1] Princeton University
[2] Michigan State University
[3] Rice University

Abstract: More than 20 years ago, researchers proposed that individual differences in performance in such domains as music, sports, and games largely reflect individual differences in amount of *deliberate practice*, which was defined as engagement in structured activities created specifically to improve performance in a domain. This view is a frequent topic of popular-science writing—but is it supported by empirical evidence? To answer this question, we conducted a meta-analysis covering all major domains in which deliberate practice has been investigated. We found that deliberate practice explained 26% of the variance in performance for games, 21% for music, 18% for sports, 4% for education, and less than 1% for professions. We conclude that deliberate practice is important, but not as important as has been argued.

Keywords: deliberate practice; talent development; meta-analysis; human performance; skill acquisition; expertise; open data

Why do so few people who take up an instrument such as the violin, a sport such as golf, or a game such as chess ever reach an expert level of performance? This question is a topic of a long-running debate in psychology. There are two classical views. One is that experts are "born"—that training is necessary to reach a high level of performance, but innate ability limits the ultimate level of performance a person can achieve. Galton (1869), the founder of behavioral genetics, argued for this position on the basis of his finding that eminence in science, music, art, sports, and other domains tends to run in families. The opposing view is that experts are "made"—that either talent does not exist or its effects on performance are overshadowed by the effect of training. Watson (1930), the founder of behaviorism, captured this view when he stated that "practicing more intensively than others . . . is probably the most reasonable explanation we have today not only for success in any line, but even for genius" (p. 212).

More recently, in the spirit of Watson, Ericsson, Krampe, and Tesch-Römer (1993) proposed their influential *deliberate-practice view* of expert performance. This view holds that expert performance largely reflects accumulated amount of deliberate practice, which Ericsson et al. defined as engagement in structured activities created specifically to improve performance in a domain. In two studies, Ericsson et al. recruited musicians with different levels of accomplishment and asked them to estimate the amount of deliberate practice they had engaged in per week for each year of their musical careers. On average, cumulative amount of deliberate practice was much higher for the most-accomplished groups of musicians than for the less-accomplished groups. For example, at age 20, the average for the "best" violinists was more than 10,000 hr, whereas the averages were about 7,800 hr for the "good" violinists and about 4,600 hr for the least-accomplished group.

Ericsson et al. (1993) concluded that "high levels of deliberate practice are necessary to attain expert level performance" and added, "Our theoretical framework can *also provide a sufficient account* [emphasis added] of the major facts about the nature and scarcity of exceptional performance. Our account does not depend on scarcity of innate ability (talent) . . ." (p. 392). They continued, "We argue that the differences between expert performers and normal adults reflect a life-long period of deliberate effort to improve performance in a specific domain" (p. 400). Ericsson (2007) reiterated this perspective when he claimed that "the distinctive characteristics of elite performers are adaptations to extended and intense practice activities that selectively activate dormant genes that all healthy children's DNA contain[s]" (p. 4).

The deliberate-practice view has inspired a great deal of interest in expert performance. A Google Scholar search in April 2014 showed that the article by Ericsson et al. (1993) has been cited more than 4,200 times (http://scholar.google.com/scholar?cites=11519303805153777449&as_sdt=20000005&sciodt=0,21&hl=en), and their research has been discussed in a number of popular books, including Gladwell's (2008) *Outliers*, Levitt and Dubner's (2009) *SuperFreakonomics*, and Colvin's (2008) *Talent Is Overrated*. Ericsson et al.'s findings were also the inspiration for what Gladwell termed the "10,000-hour rule"—the idea that it takes 10,000 hr of practice to become an expert.

At the same time, the deliberate-practice view has been sharply criticized in the scientific literature. Gardner (1995) commented that the view requires a "blindness . . . to decades of psychological theorizing" (p. 802), and Sternberg (1996) observed that "deliberate practice may be correlated with success because it is a proxy for ability: We stop doing what we do not do well and feel unrewarded for" (p. 350). Anderson (2000) stated that "Ericsson and Krampe's research does not really establish the case that a great deal of practice is sufficient for great talent" (p. 324), and Marcus (2012) concluded that "it would be a logical error to infer from the

importance of practice that talent is somehow irrelevant, as if the two were in mutual opposition" (p. 94).

Furthermore, although deliberate practice is important, growing evidence indicates that it is not as important as Ericsson and colleagues (Ericsson, 2007; Ericsson et al., 1993; Ericsson & Moxley, 2012) have argued. Gobet and Campitelli (2007) found a large amount of variability in total amount of deliberate practice even among master-level chess players—from slightly more than 3,000 hr to more than 23,000 hr. In a recent reanalysis of previous findings, Hambrick et al. (2014) found that deliberate practice accounted for about one third of the reliable variance in performance in chess and music. Thus, in these domains, a large proportion of the variance in performance is explainable by factors other than deliberate practice.

The Current Meta-Analysis

Our meta-analysis is a broad investigation of studies relevant to the deliberate-practice view. It is the first formal meta-analysis of the relationship between deliberate practice and human performance, and we cover all major domains in which this relationship has been studied: music, games, sports, professions, and education.

Our first goal was to estimate the overall correlation between amount of deliberate practice and performance. Ericsson and his colleagues have based their conclusions about the importance of deliberate practice on findings with measures reflecting the accumulated amount (i.e., number of hours) of deliberate practice (e.g., Duffy, Baluch, & Ericsson, 2004; Ericsson et al., 1993; Lehmann & Ericsson, 1996; Tuffiash, Roring, & Ericsson, 2007). Thus, we sought to answer a specific question: How much of the total variance in performance is explained by the accumulated amount of deliberate practice?

Our second goal was to investigate factors that might moderate the relationship between deliberate practice and performance. The first set of factors, which we term *theoretical moderators*, included domain (music, games, sports, professions, or education[1]) and predictability of the task environment (i.e., the degree to which the task environment can change while the performer is planning and executing an action and the range of possible actions). There were three levels of predictability—low, medium, and high. An example of an activity with a low-predictability environment was handling an aviation emergency; an example of an activity with a moderate-predictability environment was the sport of fencing; and an example of an activity with a high-predictability environment was running. We made no prediction about how the strength of the relationship between deliberate practice and performance would vary across domains. However, we did predict that this relationship would generally be more positive for high-predictability activities than for low-predictability activities, on the basis of findings that effects of training on performance are stronger when the task environment is more predictable (e.g., Ackerman, Kanfer, & Goff, 1995; Schneider & Fisk, 1982).

The second set of factors, which we term *methodological moderators*, included (a) the method used to assess deliberate practice—retrospective questionnaire, retrospective interview, or log—and (b) the method used to assess performance—expert rating of performance, standardized objective measure of performance (e.g., chess rating), group membership (e.g., amateur vs. professional), or performance on an objectively scored laboratory task. When a retrospective method is used to assess deliberate practice (questionnaire or interview), participants are asked to recall and estimate their past engagement in deliberate practice. By contrast, when the log method is used, deliberate practice is recorded on an ongoing basis, either

by the participant in a diary or by a computer. Given that people do not have perfect memory for the past, particularly the distant past, the log method presumably yields more accurate (valid) estimates of deliberate practice than retrospective methods do. Therefore, we wanted to determine whether the relationship between deliberate practice and performance differed for the log method and for the retrospective methods.

Method

We designed the meta-analysis and report the results in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (Moher, Liberati, Tetzlaff, Altman, & The PRISMA Group, 2009).

Inclusion criteria, literature search, and coding

The criteria for including a study in the meta-analysis were as follows:

- A measure of accumulated amount (e.g., number of hours) of one or more activities interpretable as deliberate practice (henceforth, *deliberate practice*) was collected, and the study report referred to at least one publication on deliberate practice by Ericsson and his colleagues.[2]
- A measure of performance reflecting level of skill in the particular domain was collected.
- An effect size reflecting the relationship between accumulated amount of deliberate practice and performance was reported, or information needed to compute this effect size could be obtained from the author(s) of the study.
- The methods and results were in English.
- The participants were human.

We did not exclude studies on the basis of participants' age or skill level.

To identify studies meeting these criteria, we systematically searched for relevant published and unpublished articles in psychology, education, sports science, medicine, and other disciplines through March 24, 2014 (for a flowchart designed according to the PRISMA specifications, see Fig. 1). We also e-mailed authors of articles on deliberate practice and requested information relevant to our meta-analysis that was not accessible (e.g., unpublished data), and we asked that they forward the e-mail to colleagues who might have conducted relevant studies.

Our search and e-mail request yielded 9,331 potentially relevant articles. After examining these articles and discarding irrelevant ones (e.g., literature reviews, commentaries), we identified 88 studies that met all the inclusion criteria. We coded each study and the measures collected in it for reference information, methodological characteristics, and results (the data file is openly available at https://osf.io/rhfsk). These studies included 111 independent samples, with 157 effect sizes and a total sample size of 11,135 participants. For a list of studies included in the meta-analysis, see the Supplemental Method and Results in the Supplemental Material available online. For additional characteristics of the meta-analysis, see Table 1.

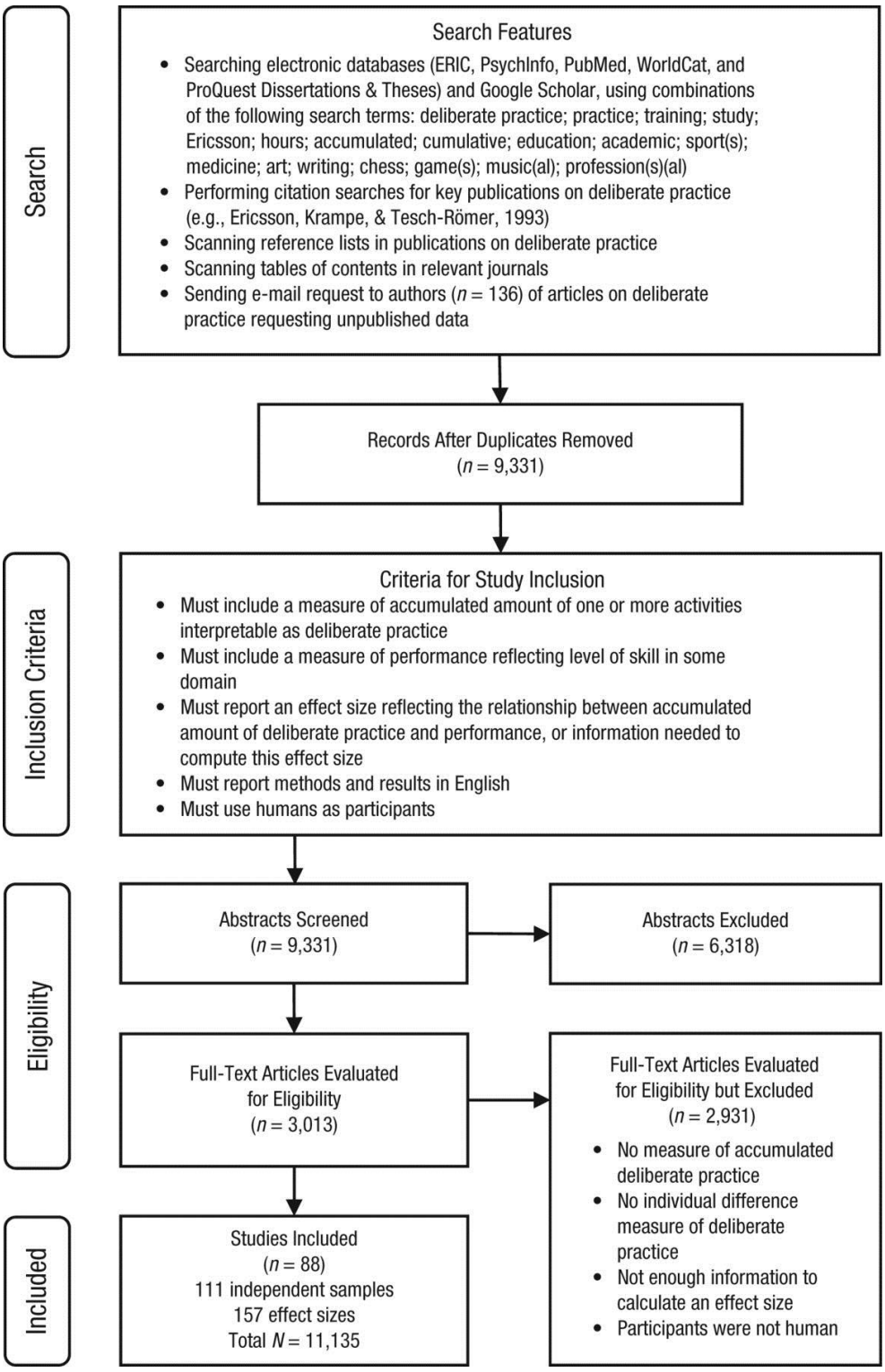

Fig. 1. Flow diagram of the literature search and study coding.

Table 1. Descriptive Characteristics of the Meta-Analysis

Study characteristic	Number of effect sizes	Number of participants
Domain		
Music	28	1,259
Games	11	1,291
Sports	60	2,633
Professions	7	321
Education	51	5,631
Method used to estimate deliberate practice hours		
Interview	36	1,238
Questionnaire	96	8,233
Log	25	1,664
Method used to estimate performance[a]		
Standardized objective measure	73	7,275
Laboratory measure	20	473
Group membership	35	2,266
Expert rating	29	1,307
Publication status		
Published	128	10,155
Unpublished	29	980
Total	157	11,135

[a]For this characteristic, the number of participants does not add up to 11,135 because some samples contributed to multiple types of effects.

The first and second authors coded each effect for moderator variables, and then two individuals with no knowledge of the effect sizes provided separate sets of coding. As indexed by Cohen's kappa for the categorical variables and Spearman's rho for the quantitative variable, interrater agreement among the independent raters and agreement between these individuals' ratings and the authors' ratings were generally high—domain: κs = .99–1.00; predictability of the task environment: ρs = .89–.96; method used to assess deliberate practice: κs = .91–.98; and method used to assess performance: κs = .78–.83. The authors resolved any discrepancies.

Effect sizes

The meta-analysis used the correlation between accumulated amount of deliberate practice and performance as the measure of effect size. For most studies, the authors reported a

correlation coefficient;[3] for studies in which the authors reported group-level comparisons (e.g., professional vs. amateur musicians), we converted standardized mean differences (Cohen's *d*s) to biserial correlations (r_b s; Becker, 1986; Hunter & Schmidt, 1990).

Meta-analytic procedure

The meta-analysis involved four steps. The first step was to obtain correlations between time spent in one or more activities interpretable as deliberate practice and performance, along with their sampling error variances. The second step was to search for extreme values. One effect size exceeded 1.0 (r = 1.15); we judged this effect size to be invalid and deleted it. There also were four outliers—effect sizes whose residuals had z scores of 3 or greater (rs = .91, .90, .90, and .84); we Winsorized these values to z scores equaling 2.99 (rs = .83, .83, .84, and .83, respectively). The third step was to estimate overall effects and heterogeneity in the effect sizes using random-effects meta-analysis modeling, and then to test whether some of the heterogeneity was predictable from moderator variables using mixed-effects meta-analysis modeling. The final step was to perform publication-bias analyses. We used the Comprehensive Meta Analysis (Version 2; Biostat, Englewood, NJ) software package to conduct the meta-analyses and publication-bias analyses. (See also Methodological Details and Screen Shots of Results, Figs. S3−S16, in the Supplemental Method and Results in the Supplemental Material.)

Results

Figure 2 shows that nearly all correlations between deliberate practice and performance were positive: High levels of deliberate practice were associated with high levels of performance. Of the small number of negative correlations (10 of 157), only 2 (< 1.5% of all correlations) were statistically significant ($p < .05$).

The meta-analytic average correlation between deliberate practice and performance was .35, 95% confidence interval (CI) = [.30, .39], which indicates that deliberate practice explained 12% of the variance in performance, 95% CI = [9%, 15%]; thus, 88% of the variance was unexplained. However, as indicated by the I^2 statistic, which specifies the percentage of the between-study variability in effect sizes that is due to heterogeneity rather than random error, there was a high degree of heterogeneity in the effect sizes, I^2 = 84.90. We investigated the source of this heterogeneity through the moderator analyses reported next.

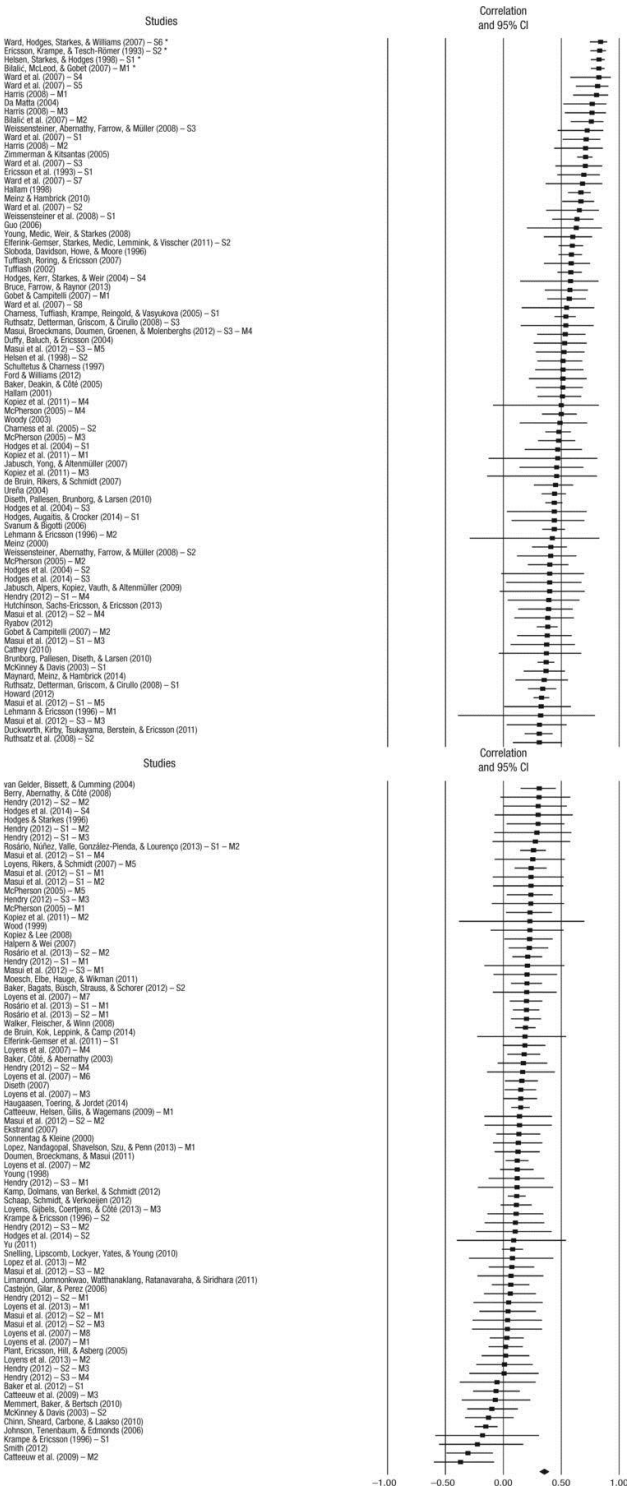

Fig. 2. Correlations between deliberate practice and performance. Correlations (squares) and 95% confidence intervals (CIs; lines) are displayed for all effects entered into the meta-analysis. The diamond on the bottom row represents the meta-analytically weighted mean correlation. Multiple measures were adjusted for dependency (see also Methodological Details in the Supplemental Method and Results in the Supplemental Material). Asterisks identify adjusted (Winsorized) outliers. For studies with multiple independent samples, the result for each sample

(S1, S2, etc.) is reported separately. Similarly, for studies with multiple performance measures, the result for each measure (M1, M2, etc.) is reported separately.

Moderator analyses

Theoretical moderators. Domain was a statistically significant moderator, $Q(4) = 49.09$, $p < .001$. Percentage of variance in performance explained by deliberate practice was 26% for games ($\bar{r} = .51$, $p < .001$), 21% for music ($\bar{r} = .46$, $p < .001$), 18% for sports ($\bar{r} = .42$, $p < .001$), 4% for education ($\bar{r} = .21$, $p < .001$), and less than 1% for professions ($\bar{r} = .05$, $p = .62$; see Fig. 3).

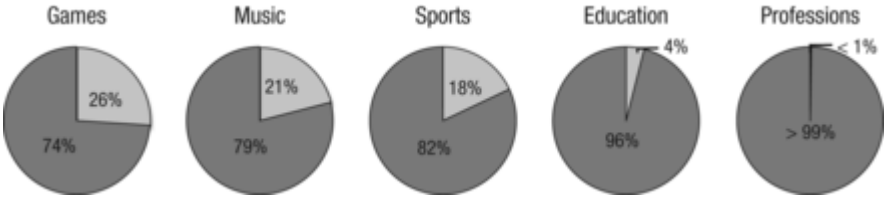

Fig. 3. Percentage of variance in performance explained (light gray) and not explained (dark gray) by deliberate practice within each domain studied. Percentage of variance explained is equal to $r^2 \times 100$.

Predictability of the task environment was also a statistically significant moderator, $Q(1) = 20.49$, $b = 0.14$, $T^2 = .05$, $p < .001$. As hypothesized, the percentage of variance in performance explained by deliberate practice was largest (24%) for activities high in predictability ($\bar{r} = .49$), intermediate (12%) for activities moderate in predictability ($\bar{r} = .35$), and smallest (4%) for activities low in predictability ($\bar{r} = .21$; see also Fig. S1 in the Supplemental Method and Results in the Supplemental Material).

Methodological moderators. The method used to assess deliberate practice was a statistically significant moderator, $Q(2) = 16.19$, $p < .001$. The percentage of variance in performance explained by deliberate practice was 20% for studies that used a retrospective interview ($\bar{r} = .45$, $p < .001$), 12% for studies that used a retrospective questionnaire ($\bar{r} = .34$, $p < .001$), and 5% for studies that used a log method ($\bar{r} = .22$, $p < .001$).[4]

The method used to assess performance was also a statistically significant moderator, $Q(3) = 14.41$, $p = .002$. The percentage of variance in performance explained by deliberate practice was 26% for studies that used group membership ($\bar{r} = .51$, $p < .001$), 14% for studies that used laboratory tasks ($\bar{r} = .37$, $p < .001$), 9% for studies that used expert ratings ($\bar{r} = .30$, $p < .001$), and 8% for studies that used standardized objective scoring measures ($\bar{r} = .28$, $p < .001$).

Additional meta-analytic models

We ran three additional models. The first model excluded the 38 effect sizes for team sports, leaving 119 effect sizes (games: 11, music: 28, individual sports: 22, education: 51, professions: 7). We ran this model because interpretation of correlations between deliberate practice and performance in team sports is complicated by the fact that an individual's performance is not independent of the team's performance (Hutchinson, Sachs-Ericsson, & Ericsson, 2013). The overall percentage of variance explained by deliberate practice was 11% in this model (games: 26%, music: 21%, sports: 19%, and education: 4%, all *p*s < .001; professions: < 1%, $p = .62$).

The second model included only the 59 effect sizes for solitary deliberate practice (games: 6; music: 9; sports: 14; education: 30; professions: 0). We tested this model to address the question of whether deliberate practice must be performed in isolation to be maximally effective (Charness, Tuffiash, Krampe, Reingold, & Vasyukova, 2005; Ericsson et al., 1993). The overall percentage of variance explained by deliberate practice was 11% in this model (games: 23%; music: 23%; sports: 22%; and education: 3%; all *p*s < .001), which indicates that solitary deliberate practice is not a stronger predictor of performance than deliberate practice with other people.

The third model included only the 53 effect sizes for solitary deliberate practice available after excluding effect sizes for team sports (games: 6; music: 9; individual sports: 8; education: 30). The overall percentage of variance explained by deliberate practice was 10% in this model (games: 23%; music: 23%; sports: 28%; and education: 3%; all *p*s < .001).

Thus, results of the additional analyses were similar and consistent with the overall analysis, indicating that deliberate practice explained a considerable amount of the variance in performance, but a large amount of the variance remains unexplained.

Publication-bias analyses

We conducted publication-bias analyses to investigate whether null or weak results have been systematically suppressed from publication in the deliberate-practice literature and whether there were effect sizes missing from our meta-analysis because of publication bias. We first inspected a funnel plot depicting the relationship between standard error and effect size; it was approximately symmetrical, suggesting that smaller-sample studies with weak effect sizes were not missing from our meta-analysis (see Fig. S2 in Additional Publication-Bias Analyses in the Supplemental Method and Results in the Supplemental Material). A trim-and-fill analysis (Duval & Tweedie, 2000a, 2000b) confirmed this, indicating that no effects were missing because of publication bias.

General Discussion

More than 20 years ago, Ericsson et al. (1993) argued that "individual differences in ultimate performance can largely be accounted for by differential amounts of past and current levels of practice" (p. 392). Ericsson and Moxley (2012) reiterated this claim, stating that "the concept of deliberate practice can account for the large individual differences between experts and novices" (p. 145). The results of this meta-analysis do not support these strong claims. Regardless of domain, a large amount of variance in performance is not explained by deliberate practice and is potentially explainable by other factors. We conclude that amount of deliberate practice—although unquestionably important as a predictor of individual differences in performance from both a statistical and a practical perspective—is not as important as Ericsson and his colleagues have argued.

Moderator analyses revealed that the strength of the relationship between deliberate practice and performance varied by domain. In terms of percentage of variance in performance explained, the effect of deliberate practice was strong for games (26%), music (21%), and sports (18%), and much weaker for education (4%) and professions (< 1% and not statistically significant). Why were the effect sizes for education and professions so much smaller? One possibility is that deliberate practice is less well defined in these domains. It could also be that in

some of the studies, participants differed in amount of prestudy expertise (e.g., amount of domain knowledge before taking an academic course or accepting a job) and thus in the amount of deliberate practice they needed to achieve a given level of performance.

Moderator analyses further revealed that the effect of deliberate practice on performance tended to be larger for activities that are highly predictable (e.g., running) than for activities that are less predictable (e.g., handling an aviation emergency), as we hypothesized. Furthermore, the effect of deliberate practice on performance was stronger for studies that used retrospective methods to elicit estimates of deliberate practice than for those that used a log method. In fact, for studies using the log method, which presumably yields more valid estimates than retrospective methods do, deliberate practice accounted for only 5% of the variance in performance. This finding suggests that the use of what Ericsson (2014) termed a "high-fidelity" (p. 13) approach to assessing deliberate practice (e.g., video monitoring) might reveal that the relationship between deliberate practice and performance is weaker than the results of this meta-analysis indicate. Finally, the relationship between deliberate practice and performance was weaker for studies that used a standardized objective measure of performance (e.g., chess rating) than for studies that used group membership as the measure of performance.

We did not correct individual effect sizes for the attenuating effect of measurement error (i.e., measurement unreliability), because very few studies in the meta-analysis reported a reliability estimate for both deliberate practice and performance. However, measures of both deliberate practice and performance are typically found to have acceptable or better reliability ($\geq$ .70). For example, Tuffiash et al. (2007) stated that test-retest reliabilities for self-report practice estimates in sports and music are typically at or above .80, and Hambrick et al. (2014) found reliability of .91 for chess ratings. Furthermore, the percentage of variance in performance explained by deliberate practice is smaller than the percentage of variance not explained by deliberate practice[5] across a wide range of reliability assumptions (see Table S1 in the Supplemental Method and Results in the Supplemental Material). For example, if it is assumed that reliability of both deliberate practice and performance is .80, the mean overall correlation between deliberate practice and performance is .43 after correction for unreliability. This correlation indicates that deliberate practice accounts for 19% of the reliable variance and that 81% of the reliable variance is potentially explainable by other factors; corresponding percentages of variance explained are 41% for games, 33% for music, 28% for sports, 7% for education, and less than 1% for professions.

What explains the variance in performance that deliberate practice does not explain? There are probably many factors. One may be the age at which a person starts serious involvement in a domain. Ericsson et al. (1993) argued that any performance advantage associated with starting age simply reflects the fact that a person who starts at a young age has more time to accumulate deliberate practice than a person who starts at a later age. However, Gobet and Campitelli (2007) and Howard (2012) found that starting age negatively predicted chess rating even after statistically controlling for deliberate practice. This evidence suggests that there may be an optimal developmental period for acquiring complex skills, as there seems to be for acquiring language (Lenneberg, 1967).

Research suggests that general intelligence and more specific abilities may also explain some of the variance in performance that deliberate practice does not. General intelligence (Hunt, 2011; Jensen, 1998)—which is highly stable and substantially heritable (Plomin, DeFries, McClearn, & McGuffin, 2008)—positively predicts performance in a wide range of domains, including music (Shuter, 1968), chess (Grabner, Stern, & Neubauer, 2007), academics (e.g.,

Brody, 1997; Laidra, Pullmann, & Allik, 2007), and virtually any occupation (Schmidt & Hunter, 1998, 2004). Working memory capacity—the ability to maintain information in the focus of attention (Engle, 2002)—is an example of a specific ability that may predict performance differences. Meinz and Hambrick (2010) found that working memory capacity positively predicted pianists' performance in a sight-reading task, above and beyond deliberate practice. There was no significant interaction between deliberate practice and working memory capacity, which indicates that working memory capacity was as important a predictor of performance for beginning pianists as it was for pianists who had engaged in thousands of hours of deliberate practice.

Conclusion

Ericsson and his colleagues' (1993) deliberate-practice view has generated a great deal of interest in expert performance, but their claim that individual differences in performance are largely accounted for by individual differences in amount of deliberate practice is not supported by the available empirical evidence. An important goal for future research on expert performance is to draw on existing theories of individual differences (e.g., Ackerman, 1987; Gagné, 2009; Schmidt, 2014; Simonton, 2014) to identify basic abilities and other individual difference factors that explain variance in performance and to estimate their importance as predictor variables relative to deliberate practice. Another important goal is to continue to investigate how and when task and situational factors such as task predictability moderate the impact of deliberate practice and other individual difference factors on performance. Research aimed at addressing these goals will shed new light on the underpinnings of expert performance.

Acknowledgments: We thank Neil Charness, Donald Chinn, Sarah Doumen, Joakim Ekstrand, Marije Elferink-Gemser, Mathias Haugaasen, Nicola Hodges, Rachelle Kamp, Thirayoot Limanond, Enrique Lopez, Gary McPherson, Joanne Ruthsatz, Igor Ryabov, Lydia Schaap, Steve Schultetus, Judy Sheard, Soren Svanum, and Marjorie Torenbeek for providing unpublished data, additional data and results associated with publications, and clarification of methods or results.

Article Notes

Declaration of Conflicting Interests: The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Supplemental Material: Additional supporting information may be found at http://pss.sagepub.com/content/by/supplemental-data

Notes

1. Studies of professional athletes were included in the sports category, and studies of professional musicians were included in the music category; the professions category included professions not captured by the other domains: computer programming, military aircraft piloting, soccer refereeing, and insurance selling. Studies included in the education category were primarily studies of university students in which the achievement outcome was a course grade or semester grade point average. See the performance measure descriptions in Column N of the Open Data file available at https://osf.io/rhfsk. We classified ballet as a sport because it is a highly physical activity and has similarities to sports such as gymnastics and figure skating.
2. For studies in which the total amount of time that participants had to accumulate deliberate practice was a constant (e.g., a college semester), we were able to use weekly amount of deliberate practice as a measure of accumulated amount of deliberate practice, given that this variable and accumulated amount of deliberate practice would necessarily have the same correlation with performance. The focus of this meta-analysis was on the relationship between individual differences in accumulated deliberate practice and performance. We did not include studies that experimentally manipulated training and then compared trained and untrained individuals.
3. We reversed the sign of the correlation when appropriate before analyzing the data. For instance, negative correlations between deliberate practice and race times in sports indicate that more deliberate practice is associated with lower (faster) race times (i.e., more deliberate practice is associated with better performance).
4. Whether or not the researchers performed a transformation (e.g., log) on the deliberate-practice variable before performing analyses was not a statistically significant moderator of the relationship between deliberate practice and performance, $Q(1) = 1.77$, $p = .18$.
5. The standard formula for correcting a correlation for measurement unreliability is $\hat{r} = r_{xy} / (r_{xx}\, r_{yy})^{1/2}$, where r_{xx} and r_{yy} are reliability coefficients for x and y, respectively (Schmidt & Hunter, 1996).

References

Ackerman P. L. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. *Psychological Bulletin*, 102, 3–27. doi:10.1037/0033-2909.102.1.3

Ackerman P. L., Kanfer R.,Goff M. (1995). Cognitive and noncognitive determinants and consequences of complex skill acquisition. *Journal of Experimental Psychology: Applied*, 1, 270–304. doi:10.1037/1076-898X.1.4.27

Anderson J. R.(2000). *Learning and memory: An integrated approach* (2nd ed.). New York, NY: John Wiley & Sons.

Becker G.(1986). Correcting the point-biserial correlation for attenuation owing to unequal sample size. *The Journal of Experimental Education*, 55, 5–8.

Brody N.(1997). Intelligence, schooling, and society. *American Psychologist*, 52, 1046–1050. doi:10.1037/0003-066X.52.10.1046

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. doi:10.1002/acp.1106

Colvin G.(2008). *Talent is overrated: What really separates world-class performers from everybody else*. New York, NY: Penguin Group.

Duffy L. J., Baluch B., Ericsson K. A. (2004). Dart performance as a function of facets of practice amongst professional and amateur men and women players. *International Journal of Sport Psychology*, 35, 232–245.

Duval S. J., Tweedie R. L.(2000a). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, 95, 89–98. doi:10.1080/01621459.2000.10473905

Duval S. J.,Tweedie R. L.(2000b). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56, 455–463. doi:10.1111/j.0006-341X.2000.00455.x

Engle R. W.(2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, 11, 19–23. doi:10.1111/1467-8721.00160

Ericsson K. A.(2007). Deliberate practice and the modifiability of the body and mind: Toward a science of the structure and acquisition of expert and elite performance. *International Journal of Sport Psychology*, 38, 4–34.

Ericsson K. A.(2014). Why expert performance is special and cannot be extrapolated from studies of performance in the general population: A response to criticisms. *Intelligence*, 45, 81–103. doi: 0.1016/j.intell.2013.12.001

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. doi:10.1037/0033-295X.100.3.363

Ericsson K. A., Moxley J. H. (2012). The expert performance approach and deliberate practice: Some potential implications for studying creative performance in organizations. In Mumford M. D. (Ed.), *The handbook of organizational creativity* (pp. 141–167). London, England: Academic Press.

Gagné F.(2009). Talent development as seen through the differentiated model of giftedness and talent. In Balchin T., Hymer B., Matthews D. (Eds.), *The Routledge international*

companion to gifted education (pp. 32–41). Abingdon, Oxfordshire, England: Routledge.
Galton F.(1869). *Hereditary genius*. London, England: Macmillan.
Gardner H. (1995). "Expert performance: Its structure and acquisition": Comment. *American Psychologist*, 50, 802–803. doi:10.1037/0003-066X.50.9.802
Gladwell M.(2008). *Outliers: The story of success*. New York, NY: Little, Brown.
Gobet F., Campitelli G. (2007). The role of domain-specific practice, handedness, and starting age in chess. *Developmental Psychology*, 43, 159–172. doi:10.1037/0012-1649.43.1.159
Grabner R. H., Stern E., Neubauer A. C. (2007). Individual differences in chess expertise: A psychometric investigation. *Acta Psychologica*, 124, 398–420. doi:10.1016/j.actpsy.2006.07.008
Hambrick D. Z., Oswald F. L., Altmann E. M., Meinz E. J., Gobet F., Campitelli G. (2014). Deliberate practice: Is that all it takes to become an expert? *Intelligence*, 45, 34–45. doi:10.1016/j.intell.2013.04.001
Howard R. W.(2012). Longitudinal effects of different types of practice on the development of chess expertise. *Applied Cognitive Psychology*, 26, 359–369. doi:10.1002/acp.1834
Hunt E. (2011). *Human intelligence*. Cambridge, England: Cambridge University Press.
Hunter J. E., Schmidt F. L.(1990). Dichotomization of continuous variables: The implications for meta-analysis. *Journal of Applied Psychology*, 75, 334–339. doi:10.1037/0021-9010.75.3.334
Hutchinson C. U., Sachs-Ericsson N. J., Ericsson K. A.(2013). Generalizable aspects of the development of expertise in ballet across countries and cultures: A perspective from the expert performance approach. *High Ability Studies*, 24, 21–47. doi:10.1080/13598139.2013.780966
Jensen A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.
Laidra K., Pullmann H., Allik J. (2007). Personality and intelligence as predictors of academic achievement: A cross-sectional study from elementary to secondary school. *Personality and Individual Differences*, 42, 441–451. doi:10.1016/j.paid.2006.08.001
Lehmann A. C., Ericsson K. A. (1996). Performance without preparation: Structure and acquisition of expert sight-reading and accompanying performance. *Psychomusicology*, 15, 1–29. doi:10.1037/h0094082
Lenneberg E. H. (1967). *Biological foundations of language*. Oxford, England: Wiley.
Levitt S. D., Dubner S. J. (2009). *SuperFreakonomics: Global cooling, patriotic prostitutes, and why suicide bombers should buy life insurance*. New York, NY: HarperCollins.
Marcus G. (2012). *Guitar zero: The science of becoming musical at any age.* New York, NY: Penguin.
Meinz E. J., Hambrick D. Z. (2010). Deliberate practice is necessary but not sufficient to explain individual differences in piano sight-reading skill: The role of working memory capacity. *Psychological Science*, 21, 914–919. doi:10.1177/0956797610373933
Moher D., Liberati A., Tetzlaff J., Altman D. G., & The PRISMA Group. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA statement. *PLoS Med* 6: Article e1000097. Available at http://www.plosmedicine.org/article/info:doi/10.1371/journal.pmed.1000097
Plomin R., DeFries J. C., McClearn G. E.,McGuffin P. (2008). *Behavioral genetics* (Vol. 5). New York, NY: Worth.
Schmidt F. L. (2014). A general theoretical integrative model of individual differences in interests, abilities, personality traits, and academic and occupational achievement: A

commentary on four recent articles. *Perspectives on Psychological Science*, 9, 211–218. doi:10.1177/1745691613518074
Schmidt F. L., Hunter J. (2004). General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology*, 86, 162–173. doi:10.1037/0022-3514.86.1.162
Schmidt F. L., Hunter J. E. (1996). Measurement error in psychological research: Lessons from 26 research scenarios. *Psychological Methods*, 1, 199–223. doi:10.1037/1082-989X.1.2.199
Schmidt F. L., Hunter J. E.(1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262–274. doi:10.1037/0033-2909.124.2.262
Schneider W., Fisk A. D. (1982). Degree of consistent training: Improvements in search performance and automatic process development. *Perception & Psychophysics*, 31, 160–168. doi:10.3758/BF03206216
Shuter R. (1968). *The psychology of musical ability*. London, England: Methuen.
Simonton D. K. (2014). Creative performance, expertise acquisition, individual differences, and developmental antecedents: An integrative research agenda. *Intelligence*, 45, 66–73. doi:10.1016/j.intell.2013.04.007
Sternberg R. J. (1996). The costs of expertise. In Ericsson K. A. (Ed.), *The road to excellence: The acquisition of expert performance in the arts and sciences, sports, and games* (pp. 347–354). Mahwah, NJ: Erlbaum.
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. doi:10.1037/1076-898X.13.3.124
Watson J. B. (1930). *Behaviorism*. New York, NY: W. W. Norton.