Journal of Expertise
2019. Vol. 2(4)

ISSN 2573-2773

Right-Tail Range Restriction: A Lurking Threat to Detecting Associations between Traits and Skill among Experts

Harrison J. Kell[1] and Jonathan Wai[2]
[1]Academic to Career Research Center, Educational Testing Service, USA
[2]Department of Education Reform and Department of Psychology, University of Arkansas, USA

Correspondence: Harrison J. Kell: hkell@ets.org; Jonathan Wai: jwai@uark.edu

Abstract

It has been claimed by prominent authors that there is no relationship between differences in some human traits (e.g., cognitive ability, physical ability) and differences in skill among experts. We assert that the failure to detect such associations is often due to an extreme form of range restriction that particularly plagues research focused on expert samples: *right-tail range restriction* (RTRR). RTRR refers to a lack of representation of data from the far right segment of the normal distribution, inhibiting the observation of statistical associations. Using two example studies we demonstrate that, when RTRR is not present, relationships between differences in experts' traits and differences in their degree of skill can be observed. Based on the characteristics of these studies we make recommendations for methodological practices that can be followed to help investigators overcome RTRR and facilitate the continued development of a robust and replicable science of expertise.

Keywords

Range restriction, expertise, traits, cognitive ability, physical ability, performance, athletics, psychological attributes

Introduction

Historical interest in experts is longstanding, with a focus on extraordinary achievement across domains ranging from athletics to scientific innovation (Simonton, 1994). Interest in extraordinary achievers stems in part from the desire to understand the factors that may contribute to achieving expertise and excellence. Such factors include the practices that must be carried out in order to develop expertise as well as the psychological and physical attributes that facilitate and undergird the acquisition of expertise. Preliminary research into these factors, especially relatively stable human attributes (traits), primarily rests on observing consistent associations.[1]

Psychological attributes related to expertise have been of scientific interest for over 150 years (Quetelet, 1842); Galton's (1869) *Hereditary Genius* was the first major work treating the topic in the English-speaking world. Galton is known for advocating that cognitive ability is the primary driver of extreme achievement, but he also postulated that intense energy and persistence were necessary for great accomplishment (Simonton, 1991). Since then many other traits have been hypothesized to be associated with expertise, including situation awareness (Endsley, 2006), emotion regulation (Grabner, Stern, & Neubauer, 2007), and controlled attention (Haber & Haber, 2003).

An enduring affliction of studying associations is range restriction (Thorndike, 1949). In the expertise literature, range restriction has often been invoked in discussions of potentially causal relations between cognitive ability and expertise, as many studies have found trivial to non-existent correlations between skilled performance and ability. The absence of a strong association between the trait of cognitive ability and skilled performance has led some (e.g., Ericsson, Krampe, & Tesch-Römer, 1993) to conclude that ability plays no role in expertise development. This claim is based on the reasoning that, while correlation does not equal causation, it is usually a prerequisite for causal inference (Gower, 1997). Recently, the null association between expertise and cognitive ability has been challenged on the ground that reliance on studying experts truncates variation in ability by excluding much of the general population (McAbee, 2018). To counter this shortcoming some scholars have recommended examining variation in skill across the entire proficiency spectrum, rather than focusing solely on experts (McAbee & Oswald, 2017). Indeed, although much of the discussion of range restriction in expertise research has focused on cognitive ability, the topic is generic and applicable to practically any trait presumed to be associated with expertise—psychological or otherwise. A salient example in sport is the trivial, and sometimes even negative, correlation between professional basketball players' heights and their individual and team performance (Caruso, Fleming, & Spector, 2014; Teramoto & Cross, 2018).

We extend this discussion to considering the role that range restriction may play when examining potential relations between differences in traits and differences in skill *among experts themselves*. Whereas previous discussions have been concerned with showing that there are often substantial correlations between traits and skill when less restricted samples than experts are studied, here we are concerned with raising awareness of the fact that there may be substantial variation in some traits even among experts, that it may be related to skill differences among those experts, and discussing why this variability may be obscured by current practices in—and the challenges of—expertise research. That experts may differ substantially in some of their attributes is important for theory development and stimulating future research (Underwood, 1975), as such findings can provide a rationale for investigating whether those traits may be important for acquiring expertise and continued skill development after expert status has been achieved (Macnamara, Hambrick, & Oswald, 2014).

Examples of the belief that differences in enduring attributes among experts are unrelated to differences in experts' performance are less common than typical statements about range restriction in expertise research, given that much expertise research focuses on comparing experts to novices, rather than examining variation in the skills of experts alone (Vaci, Gula, & Bilalić, 2014). Nonetheless, perhaps because of the pervasiveness of cognitive ability tests (Kell, 2018)—and their controversial history (Gould, 1996)—it has been suggested that ability differences between experts are not related to differences in their skill: "The relationship between success and IQ works only up to a point. Once someone has reached an IQ of somewhere around 120, having additional IQ points doesn't seem to translate into any measurable real-world advantage" (Gladwell, 2008, p. 78), "A person with a 150 IQ is in theory much smarter than a person with a 120 IQ, but those additional 30 points produce little measurable benefit when it comes to lifetime success" (Brooks, 2011, pp. 284-285), and "The average IQ of scientists is certainly higher than the average IQ of the general population, but among scientists there is no correlation between IQ and scientific productivity… among those who have become professional scientists, a higher IQ doesn't seem to offer an advantage" (Ericsson & Pool, 2016, pp. 234-235).

Skepticism as to whether differences in some attributes matter among experts is not limited to cognitive ability. For instance, it has been claimed that "although aerobic capacity and upper body strength are important in basketball, levels greater than that possessed by the average college basketball player do not offer any further advantage in determining playing time" (Hoffman, Tenenbaum, Maresh, & Kraemer, 1996, p. 70), that scores on the National Football League (NFL) Scouting Combine are not "job-related," with the "job" being that of professional football player (Ledbetter, 2011, p. 2), and that soccer players' fitness scores "[do] not distinguish consistently between professionals and internationals" (le Gall, Carling, Williams, & Reilly, 2010, p. 93). More generally, questions about whether associations between various traits and performance level off (i.e., dwindle to null or nil) after a certain threshold is reached or continue even into the highest range of performance are really questions about the relation between traits and skill among experts. The tension between the two perspectives is concisely expressed in the title of an article examining ability and performance: "More-Is-Better" versus "Good-Enough" (Arneson, Sackett, & Beatty, 2011).

Our goal is to explicate a phenomenon that we believe inhibits detecting consistent associations between experts' traits and their degree of skill, an extreme form of range restriction we call RTRR. We explore the implications of RTRR for expertise research, including how overcoming it can contribute to a more robust science of highly-skilled performance and failing to overcome it may lead to illusory failures of replication. First, we define RTRR and explain how it can afflict predictors and criteria in expertise studies. Next, we summarize the results of several prior investigations to demonstrate that variation in one well-studied trait—general cognitive ability—is related to variation in experts' proficiency. We draw on these prior investigations, and the expertise literature more broadly, to make recommendations for avoiding RTRR. We then discuss the implications of these methodological practices for building a high-quality science of expertise, including demonstrating how failing to implement these practices may undermine reproducibility.

Two Types of Range Restriction in Expertise Research

Relationships between variables are often indexed by correlations, which are standardized covariances (Rodgers & Nicewander, 1988); without appropriate variation in the scores, meaningful statistical associations are difficult to observe. When the variability of one variable is truncated relative to its variability in the larger population of interest it is termed *range restriction*, which often artificially reduces the apparent association between the restricted variable and others. The implications of range restriction in expertise research have been discussed and explored previously (e.g., Ackerman, 2014). Here, we build upon these prior contributions by distinguishing between *left-hand* and *right-hand range restriction*, focusing on extreme right-hand range restriction: RTRR.

When there is a positive association between skill and some attribute and only highly skilled individuals (experts) are studied it restricts that attribute's range. Individuals scoring low(er) on the attribute will be excluded from the sample being studied, with the degree of restriction dictated by the strength of the association between the attribute and skill in the domain. Because the attribute scores of the individuals excluded will tend to be lower than those of subjects included in the expert sample we label this left-hand range restriction, as those scores will trend toward the left-hand side of a normal distribution. Likewise, if there tends to be an absence of individuals scoring high(er) on an attribute in the sample we label it right-hand range restriction (see Figure 1).

Here we are concerned with extreme right-hand range restriction, which we call RTRR. We posit that when researchers fail to detect (or find evidence for only marginal) associations between experts' traits and their degrees of skill it often may be due to 1) lacking a large sample of experts that vary in their proficiency, and 2) lacking a large sample of individuals scoring in the very top percentiles (e.g., 1%) of the attribute(s) in question. We label such cases examples of RTRR because the variability in the two variables being excluded is concentrated in the far-right segment of the normal distribution, represented by its right-tail.

Whether range restriction occurs in a sample is defined by whether variability in that sample is truncated relative to the presumed variability in the population of interest (McAbee & Oswald, 2017). Because the continuum of skilled performance reaches from those who qualify as experts in only the most basic sense to the very best performers in the world, unless a sample includes participants who meaningfully vary in their degree of expertise, that sample will be range restricted relative to the population of experts in that domain. For example, if the relationship between conscientiousness and boxing skill among experts was being studied it would not be enough to study a sample of professional boxers to avoid RTRR—the sample would also have to feature top-ranked professional boxers, including present champions and champions from the recent past. Avoiding RTRR among experts entails studying not only the very good and the great, but also the "very great" and, ideally, the greatest.

Similarly, unless a sample includes substantial numbers of experts who score in the very highest reaches of the attribute of interest, that trait's variance will be range restricted relative to its variability in the total population of experts. Because many studies of expertise suffer from at least one of these varieties of RTRR—and many suffer from both—the association between trait-like differences and differences in experts' proficiency is frequently presumed to be small to nonexistent. In the following sections we draw on data from several previously published studies to illustrate that it *can* be demonstrated that differences in experts' traits overlap with differences in their degrees of skill, when appropriate methodological characteristics are present.

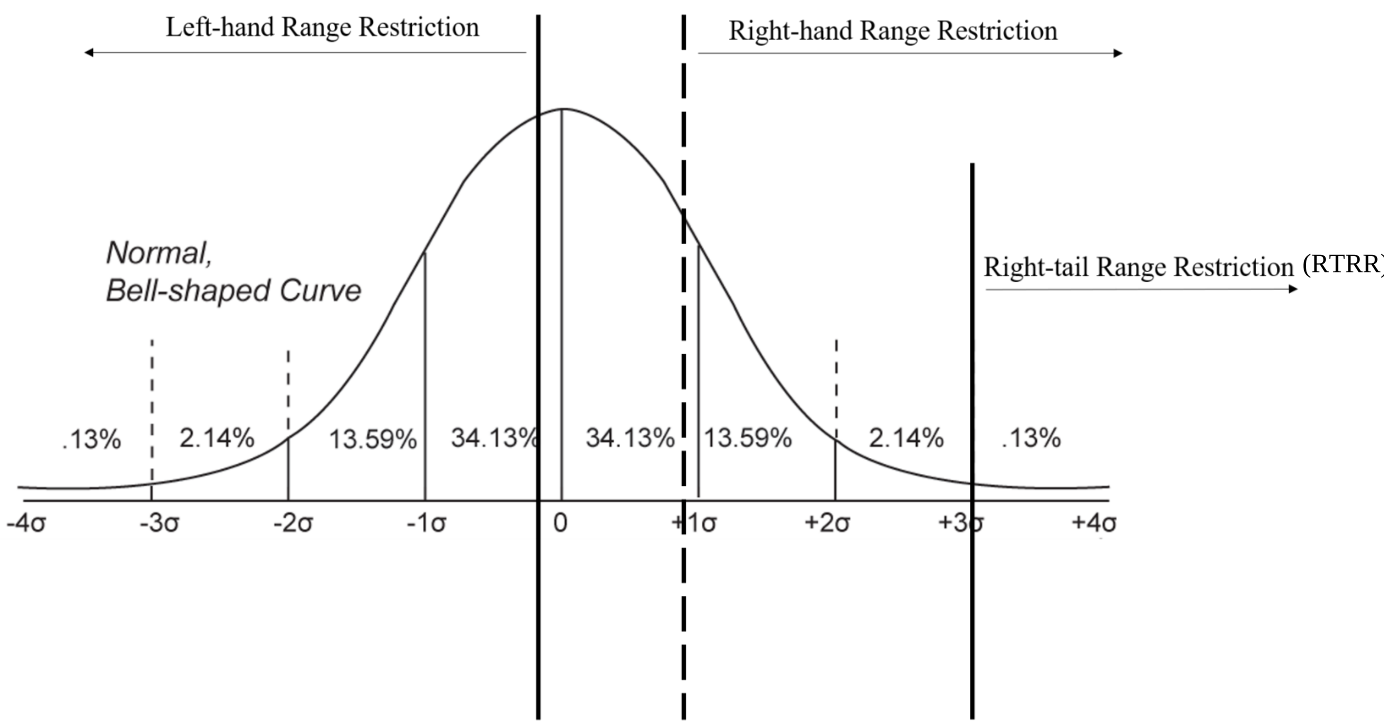

Figure 1. Visual illustration of three types of range restriction applied to the normal distribution.

Two Example Studies

There are two major ways to study the relationship between differences in human traits and differences in expertise. The first is to identify a large sample of experts demonstrating considerable variability in their achievements (e.g., national competition winners, Olympic medalists). After identifying this sample, evidence of the experts' trait of interest can be gathered, to gauge the extent to which its variability overlaps with variability in their proficiency. This approach is *criterion-centered* (Astin, 1964) in that individuals are chosen for inclusion based on their skill within a performance domain (the criterion) rather than the construct(s) hypothesized to be associated with variance in that skill. This approach is illustrated in Example Study 1 (ES1).

A second approach is to first identify a large sample of individuals who score very high on the attribute of interest *and* that demonstrates considerable variability within that high range of the trait. Such a sample might consist of, for example, participants scoring just above the cut for being considered to belong to the top 2% of a given attribute, along with participants scoring at the cuts denoting the top 1% and top 0.5%. After identifying this sample, evidence of these individuals' expertise can be gathered, to gauge the extent to which its variability overlaps with participants' trait variability. This approach is *predictor-centered* in that individuals are chosen for inclusion based on their trait standing (the hypothesized predictor) rather than the thing to be predicted, their expert performance. This approach is illustrated in Example Study 2 (ES2).

The two example studies we present focus on the relationship between general cognitive ability and expertise in various domains. Perhaps due to the heated debate about the role this trait plays in skill development (e.g., Ackerman, 2014) such research is prevalent compared to investigations of the association between other traits and skilled performance. Consequently, it is important to note that the two studies are intended to serve as examples of how differences in traits and differences in expertise can be uncovered, rather than specific lessons in the implications of ever higher levels of cognitive ability for ever higher levels of achievement. Both studies focus on data trends and patterns, as detecting empirical sequences and configurations can play an important role in scientific discovery and understanding (Bogen & Woodward, 1988; Meehl, 2004).

Study 1: Differences in Inferred Cognitive Ability Among Highly Successful Individuals

The findings comprising ES1 are drawn from studies that have investigated a variety of occupational experts worldwide (Wai, 2013; Wai & Rindermann, 2015; Wai & Lincoln, 2016). To align these samples more closely with those studied in ES2 we focus solely on those from the U.S., and specifically experts in politics and building wealth, due to the influence such individuals have in shaping modern society and culture (for better or worse). The sample of political experts is comprised of members of the U.S. House of Representatives (N=441) and Senate (N=100) (Wai, 2013). The sample of wealth acquisition experts is comprised of individuals identified by the organization Wealth-X as having a net worth of more than $30 million ("$30^+$-millionaires"; N=8,649) or a net worth of $1 billion or more ($N$=588).

The second author attempted to retrospectively estimate these experts' cognitive abilities by first ascertaining the institutions they attended for their undergraduate or graduate educations and then by identifying the institutions with the highest average standardized test scores. As standardized tests such as the SAT and ACT are strongly associated with general cognitive ability overall (Frey & Detterman, 2004), the institution the individual attended was used as a rough proxy for that person's ability level. An institution was designated as "elite" if its average standardized test score corresponded to an inferred cognitive ability level of roughly the top 1% relative to the general population, as indicated by the distributions of SAT and ACT scores. Due to the very high requirements, and unique stature, of Harvard University – even within elite institutions—it was coded independently of the other schools, as these extreme requirements corresponded to an inferred ability level considerably higher than the cut-off for the top 1%.

Table 1. Summary of Results for Example Study 1

Expertise Domain \ Ability Level	Elite School	Harvard
	Wealth	
Net worth ≥ \$30 million	.338	.090
Net worth ≥ \$1 billion	.434	.122
	Politics	
U.S. House of Representatives Members	.206	.066
U.S. Senators	.410	.120

Note. Percentages for experts in building wealth are based on data from Wai and Lincoln (2016, p. 17, Appendix A). Percentages for political experts are based on data from Wai (2013, p. 206, Table 2).

Table 1 presents the percentages of the four expert categories that attended an elite institution or Harvard; there is evidence that increasing expertise accompanies increasing ability. First, a greater percentage of billionaires attended an elite institution than did the 30^{+}-millionaires. Second, a greater percentage of billionaires attended Harvard than did the 30^{+}-millionaires. The same trend is evident in the domain of U.S. politics, where the position of senator is more prestigious and influential than that of House representative (Stewart & Reynolds, 1990). Notably, senators were twice as likely to attend an elite institution or Harvard as House representatives, supporting the case that even among experts more ability, on average, is associated with greater proficiency.

In summary, the data in Table 1 yields four comparisons of proportions, all of which concern whether there is a trend toward greater representation of very high ability individuals as the difficulty of achieving the performance standard increases (e.g., House representatives vs. senators). In all four cases, this trend is present: The proportion of individuals who attended elite schools increases across the two expertise categories. The same trend is seen in terms of the representation of people who attended Harvard.

Study 2: Differences in Expertise Among Individuals in the Top 1% of Cognitive Ability

ES2 draws from the Study of Mathematically Precocious Youth (SMPY). SMPY is a longitudinal study of cohorts of individuals largely distinguished by their very high scores on cognitive tests. Initiated in 1971 (Lubinski & Benbow, 2006), SMPY has primarily relied upon administering the SAT to young adolescents to identify those possessing precocious cognitive skills. The findings reviewed in ES2 are drawn from research conducted into the achievements of SMPY's first three cohorts (Lubinski & Benbow, 2006). Cohort 1 comprises individuals identified when they were 12 or 13 (1972–1974) and who scored at least in the top 1% of cognitive ability (SAT-M ≥ 390 or SAT-V ≥ 370). Cohort 2 consists of participants identified when they were 12 (1976–1979) and who scored at least in the top 0.5% of cognitive ability (SAT-M ≥ 500 or SAT-V ≥ 430). Cohort 3 is composed of individuals scoring in the top 0.01% (SAT-M ≥ 700 or SAT-V ≥ 630) and who were identified from 1980 to 1983, when they were 12. All three cohorts have been tracked longitudinally and periodically surveyed at multiple timepoints. The findings we summarize here are largely drawn from the cohorts' most recent follow-ups and are presented in Table 2.

Table 2. Summary of Results for Example Study 2

Expertise Criterion / Ability Level	Doctoral degree	Peer-reviewed publication (≥ 1)	NSF grant (≥ 1)	NIH grant (≥ 1)	Patent (≥ 1)
Top 1% (Cohort 1)	.24	.212	.023	.027	.054
Top 0.5% (Cohort 2)	.314	.322	.038	.03	.126
Top 0.01% (Cohort 3)	.44	.24	.06	.03	.15

Note. Percentages for Cohort 1 and 2's doctoral degrees are based on data from Benbow et al. (2000, p. 475, Table 1) and for all other accomplishments are based on Lubinski et al. (2014, pp. 2218 & 2220, Table 2). Percentages for all Cohort 3's accomplishments are based on data from Makel et al. (2016, p. 1012, Table 1).

Members of Cohorts 1 and 2 have entered midlife and have impressive accomplishments. Cohort 1's achievements are reported in the first row of Table 2 and were accomplished by the time participants were, on average, age 53 (*N*=1,383; Lubinski, Benbow, & Kell, 2014). To calibrate the rarity of some of these accomplishments, the base rate of earning a doctorate in the general U.S. population is approximately 2% (U.S. Census Bureau 2012a, 2012b). Given that possessing or working toward a doctoral degree is practically a prerequisite for having at least one peer-reviewed publication or having a National Science Foundation (NSF) or National Institutes of Health (NIH) grant, we reason that the base rates of these achievements are less than 2%. The table is centered on achievements that largely depend on having a doctoral degree because individuals with them—or preparing to earn them—are often treated as experts (e.g., Chi, Gleser, & Rees, 1982; Johnson et al., 1981; Voss, Greene, Post, & Penner, 1983). In the final column is the patent-holding rate, which has a base rate of about 1% (U.S. Patent and Trademark Office, 2011). We consider holding a patent a marker of expertise, as patent approval is contingent upon four criteria: The invention must be subject matter eligible, new, useful, and non-obvious (U.S. Patent & Trademark Office, 2015).

Table 2's second row reveals Cohort 2's (top 0.5%; *N*=592) accomplishments are even more impressive: Across all criteria Cohort 2 exceeded Cohort 1. What is noteworthy is that these individuals were, on average, five years younger than those in Cohort 1 at the time of achievement (Lubinski et al., 2014).

The final row displays the accomplishments of Cohort 3 (*N*=320; Kell, Lubinski, & Benbow, 2013; Makel, Kell, Lubinski, Putallaz, & Benbow, 2016). The percentage of people in the top 0.01% with at least one of the five expertise demonstrations exceeded those of the top 1% (Cohort 1) in all cases. Cohort 3's participants' achievements exceeded those of Cohort 2 in three cases (doctoral degrees, patents, NSF grants), equaled them in one (NIH grants), and did not exceed them in one (peer-reviewed publications). The average age of the constituents of Cohort 3 was 34 when data about their educational attainment was gathered and 38 when data about their achievements were gathered.

A smaller percentage of individuals scoring in the top 1% of ability achieved at least one of the five expertise markers than those scoring in the top 0.5% of ability (despite being an average of five years older) and the top 0.01% (despite being an average of 15 years older). A smaller percentage of individuals scoring in the top 0.5% acquired a doctoral degree, patent, or NSF grant than individuals scoring in the top 0.01% and an equal percentage held at least one NIH grant—despite being an average of 10 years older than those scoring in the top 0.01%.

The data in Table 2 yield three types of comparisons across the five categories: The top 1% versus the top 0.5%, the top 1% versus the top 0.01%, and the top 0.5% versus the top 0.01%. Across these comparisons, the overall trend is that greater general cognitive ability is associated with greater expertise – even when individuals with relatively lesser cognitive ability had additional time to acquire the markers of that expertise. The representation of the top 0.5% and 0.01% exceeded that of the top 1% across all five markers. The trend is not perfect when comparing Cohort 2 to Cohort 3, however. The same proportion of those in the top 0.5% and top 0.01% had at least one NIH grant while the top 0.5% exceeded the top 0.01% in terms of holders of a peer-reviewed publication.

Overcoming Right-Tail Range Restriction (RTRR)

Here we discuss practices for facilitating the observation of associations between experts' traits and their degree of skill. We highlight how these practices were implemented in the preceding example studies but also offer suggestions for extending those investigations' methodologies. Some of these practices were first sketched over a decade ago in the domain of scientific creativity (Park, Lubinski, & Benbow, 2008) but their continued neglect in the study of expertise suggests an expanded and more general treatment may contribute to their more widespread adoption.

Ensuring Variability in Expertise is Observable

Traveling far out into the right-tail of skill is necessary to overcome RTRR and reveal differences among experts. For example, in chess the Elo cut score to be considered a grandmaster is over 2,500 (Gobet & Campitelli, 2007)—but Elo scores extend to over 2,800 and the average rating of the current top 10 players is 2,789 (World Chess Federation, 2019). Similarly, the highest career batting average among active Major League Baseball players is .315—but the highest of all-time is Ty Cobb's .367 (MLB Advanced Media, 2019; Stump, 1996). The gaps between even the very great and the greatest can be substantial. Not accounting for these gaps obscures variability in skilled performance, leading to potentially inaccurate conclusions about the relations between differences in skill and differences in traits.

Defining and identifying experts. Before differences among experts can be observed it is necessary to define those experts. The definition of what makes an individual an expert has important implications for how experts are to be identified and, ultimately, how differences in skill among those experts might be observed and what they might even consist of (cf. Macnamara et al., 2014). Unfortunately, deciding upon who is an expert is neither easy nor straightforward (McAbee, 2018).

One approach to defining and identifying experts is closely tied to proximal observation of very high-quality performance. Ericsson (2014, p. 83) defines expert performance as "consistently superior performance on a specified set of representative tasks for a domain", meaning that whomever can reliably demonstrate superior performance in this way constitutes an expert. This definition has two key aspects: The display of highly skilled behaviors under controlled conditions and the proficiency of such behaviors far exceeding that of the general population (Ericsson, 2006). Because the conditions of recitals, tournaments, and other competitions closely approximate these two requirements, performances by individuals in such environments (e.g., athletes, chess players, dancers, musicians) are treated as valid indicators of expertise (Ericsson, 2006). In this "performance-centered approach", experts can sometimes be identified using publicly available information that indicates the relative standing of a person within the population of performers in a given domain.

A related approach is not directly tied to concrete, well-specified performances but does rely on widely-accessible available data. For example, the quality of a performance or product can be inferred from its popularity and general reception, such as how often a composer's piece is performed or recorded (Kozbelt, 2007). Past experts have also been identified according to the magnitude of their historical accomplishments (Chi, 2006) or the space devoted to their entries in major reference works (Murray, 2003). These methods are consistent with Wai's (2014b) "social recognition" approach to identifying experts. For example, Wai (2014b) reasoned that because billionaires are in the top 0.0000001% of wealth in the U.S. they should be treated as possessing extreme proficiency in acquiring wealth, making them experts in this domain (Wai & Kanaya, 2019).

Identifying experts by their inclusion in some well-defined group (e.g., "billionaire") is another viable method for selecting experts: the group membership approach (Chi, 2006; Kell & Lubinski, 2015). Here, membership in some group—often an occupational or educational category—is treated as a rough proxy for possessing the skill(s) of interest. This approach is most effective when the group constitutes a shorthand for where a person stands in a continuous distribution of performers. For example, an "expert" chess player has an Elo rating above 2,000 and an "international master" has a rating above 2,400 (Gobet & Simon, 1996).

Group membership is particularly useful for establishing expertise—or some relative ranking of skill—when a concrete standard must be attained in order to gain entry to a group. For example, a person cannot earn a doctoral degree without an extended period of knowledge and skill development and after being awarded that degree by individuals who themselves are considered experts (faculty members). Similarly, in the U.S. an individual cannot be a Certified Public Accountant without having passed the Uniform Certified Public Accountant Examination.

Revealing differences among experts. Regardless of how experts are defined and identified there is ample evidence for important variability in their performance in many domains—variability that studies' methodologies sometimes cannot reveal. Uncovering differences among experts is challenging, especially because even when differences are present the sample sizes will oftentimes be too small for those differences to be statistically significant. Nonetheless, we believe there are multiple approaches that can be taken to better account for differences among experts and partially alleviate the effects of RTRR.

When expertise is closely tied to proficiency in specific performance domains, studies often lack representation of experts whose performances exist at the extreme right-tail of the skill distribution or feature such a small number of these individuals that the central tendencies of their performances cannot statistically stabilize. For example, Grabner et al. (2007) recruited 90 chess players whose Elo ratings ranged from 1,311 to 2,387, thus including some players classified as "expert" (2,000 & above) and some as "master" (2,200-2,400). However, with a sample size of only 90—some of whom scored below expert-level—it is doubtful that stable differences across the far right-tail of the performance distribution could be detected. Further, although master-level players were included, performances of individuals scoring even further into the right-tail were not, suggesting some degree of RTRR remained. Similarly, even if expert athletes are considered those in the top 5% of the player population in a given sport (Baker, Wattie, & Schorer, 2015) and a large sample of such athletes is studied, without including players who have won high-level competitions in that sample the full breadth of athletic performance in that sport is not being represented.

When group membership is used to define expertise, different groups that might rationally be expected to differ in their expertise are often treated as homogeneous; some studies have homogenized faculty members, postdoctoral fellows, and graduate students (e.g., Chi et al., 1982; Johnson et al. 1981; Voss et al., 1983), treating all equivalently. Oftentimes information about the degree of advancement of experts that might be indicative of important differences in their knowledge and skill (e.g., assistant professor vs. full professor) is not provided. Even when experts' performances are examined separately, oftentimes the small sample sizes make only qualitative comparisons possible. When quantitative analyses *are* performed, experts' results are sometimes averaged, eliminating potentially valuable information about performance differences (e.g., Ericsson et al., 1993; Johnson et al. 1981; Judkins, Oleynikov, & Stergiou, 2009).

Quantitative, continuous data are always preferable; the distinction between a "master" and "grandmaster" in chess is ultimately the product of an arbitrary cutting-score—as is the very distinction between experts and non-experts. Nonetheless, categorical distinctions are commonplace in expertise science (Vaci et al., 2014) and oftentimes continuous scores indicating skill variability are not available. When group membership is underwritten by quantitative data (e.g., athletic performances) the goal will be to include experts spanning the entire performance continuum. When group membership is not tied to a quantitative score continuum rational choices must be made based upon findings in the relevant literatures that could suggest differences in the expertise of members of these categories. For example, a master carpenter is likely more skilled than a carpenter and a member of the U.S. Senate a more skilled politician than a state senator. Thus, "level in group" is a viable means of differentiating experts when they are defined in terms of their group membership. Consequently, in terms of group membership, progressively more difficult to achieve categories within that membership group could be specified that are (usually) contingent upon passing initial hurdles (e.g., doctoral degree to peer-reviewed publication to NSF grant) (Kell & Lubinski, 2015).

According to these recommendations, the two studies we presented could be improved upon in future research to potentially tease out finer-grained differences among experts. For example, in ES2 we presented evidence showing that individuals in the top 0.5% and top 0.01% outperformed those in the top 1% across five outcomes. However, these outcomes were examined only categorically—the percentage of each cohort that had at least one of them. Future research could expand upon this approach by investigating differences in, for example, the total number of peer-reviewed publications or patents per cohort, the average number of achievements per individual, the total amount of NSF or NIH funding per cohort, or individuals' *h*-indices (discipline-normed).

Timeframe of achievement. Finally, we emphasize that one critical element for differentiating experts is time. Although time and deliberate practice are not interchangeable, the benefits of deliberate practice cannot accrue without substantial time; age and experience or years active, although not equivalent to deliberate practice, are a precondition for engaging in it. More deliberate practice is an important determinant of differences between experts (Vaci et al., 2019) but without adequate time the differences in skill that manifest due to additional practice cannot occur. Accounting for the element of time is also critical because the major accomplishments that differentiate experts often emerge only after very long periods; differences between experts (and their relationship with trait differences) will not appear if assessment occurs too early.

For example, members of Cohort 2 held at least one NSF grant at a rate of over 1.5 times that of Cohort 1 while Cohort 3 constituents held an NSF grant at over 2.5 times the rate of Cohort 1. However, it takes a long time to develop the knowledge and skills needed to secure NSF grants. If the three cohorts' achievements had been assessed only at, say, age 28—when many of them were likely just earning their doctoral degrees—these larger differences would not have appeared because few to no members of the cohorts would have yet won any NSF grants. In turn, this would make it appear that, above the cut for the top 1% of ability, more ability is unrelated to obtaining NSF grants.

The timeframe of achievement is also important among expert athletes. Lyons, Hoffman, Michel, and Williams (2011) examined the association between NFL Combine scores and players' performance across four seasons. Their decision to include four years of performance data was based on the fact that NFL players' average careers lasted approximately four seasons. However, limiting criterion data to a short timeframe excludes the performances of players who have had very long careers—and long careers are characteristic of individuals considered to be the best players. According to a ranking of the greatest NFL players, the average number of years the top 10 players of all time were active is 15 (DeMeyer, Bromberg, & Hindle, 2019). Only studying a short window of experts' careers can restrict variability in the available data and exclude the accomplishments of individuals who fall far out into the right-tail of the skill distribution—or not give active experts time to accumulate the achievements that eventually distinguish them as the "best of the best".

Ensuring Variability in Trait Scores Is Observable

Just as it is necessary to travel further and further out into the right-tail of skilled performance to reveal differences among experts, so too is it necessary to do so to reveal variability in the trait being studied. This can be challenging, as many assessments may not possess enough items of sufficient difficulty to adequately differentiate individuals scoring within the upper range of a trait's distribution (Gross, 2002). In cognitive ability testing, for example, 1,700 individuals in the high school class of 2017 earned a perfect SAT score and over 2,000 did so in the high school class of 2018 (College Board, 2017, 2018a, 2018b). Although these test-takers all earned identical scores on the SAT it does not mean they possess identical degrees of cognitive ability, only that the test was not difficult enough to make fine-grained distinctions among them. As a consequence, studies relying on SAT scores earned by typical test-takers to estimate cognitive ability will likely suffer from RTRR, as those scores will not distinguish among individuals scoring "merely" in the top 1% versus those scoring in the top 0.01%. Such circumstances could artifactually trivialize differences among experts in terms of their cognitive abilities. For example, if 57% of individuals with a net worth equal to or greater than $30 million registered perfect SAT scores versus 59% of those with a net worth of $1 billion or more, it would indeed appear that differences in cognitive ability do not have "any measurable real-world" relationship (cf. Gladwell, 2008, p. 78) with differences in prowess in acquiring wealth among those already considered experts in acquisitiveness.

The same phenomenon can occur when the physical ability of prospective professional football players is assessed by the NFL Combine (Lyons et al., 2011). Nearly all individuals who participate in the Combine have played collegiate football and thus might already be considered experts. Within this extremely select group an even more physically able group is recruited to play in the NFL. Amongst this group of expert football players there is little variability in indicators of their physical ability, as the Combine scores are not precise enough to distinguish players who are far out in the right-tail of physical ability from those *very* far out in the right-tail. Accordingly, correlations between average performance in the NFL and Combine scores can be inconsequential, even negative (Lyons et al., 2011). Similarly, perhaps highly successful people in jobs that tend to require prodigious socializing (e.g., event planning, fundraising, sales) all score at or near the ceiling of traditional measures of extraversion, making it difficult to distinguish amongst them according to this trait and, accordingly, limiting investigators' ability to detect associations between extraversion and skill amongst experts in those occupations.

Both example studies demonstrate methods for overcoming the limitation of the lack of precision of many cognitive tests when estimating the upper boundaries of ability. In ES1 attendees of Harvard were separated from those attending other elite schools, due to Harvard's more stringent admissions requirements. This method was inevitably imprecise as it relied on retrospectively inferring individuals' ability levels from institution-level data. Nonetheless, it allowed for the separation of individuals within the top 1% of cognitive ability which, in turn, allowed for the discovery that the percentage of individuals scoring higher in ability tended to increase with the degree of expertise manifested. Without the separation of individuals at the top 1% (elite institutions) and beyond the top 1% (Harvard) this trend could not have been uncovered, as there would solely have been a single ability band to examine (i.e., only elite institutions).

In ES2 investigators administered the SAT, a measure typically taken by 16- to 18-year-olds intending to enroll in college, to 12- and 13-year-olds. This approach (above-level testing) is capable of revealing differences in cognitive ability that are not perceptible when test-takers reach later adolescence, as the SAT will be much more difficult for younger test-takers. This difficulty allows for differences in the top 1% of ability to be more easily observed, especially as younger test-takers will be forced to rely more on reasoning than knowledge to solve the problems, given that they will not yet have been exposed to much of the test content in their regular schoolwork (Benbow, 1988).

Implications for Robustness and Replicability

The two example studies presented evidence suggesting that meaningful differences in skills exist among experts (at least in some fields), that these differences can be accompanied by differences in human traits, and that discovering these differences can be facilitated by adopting a variety of methods to overcome RTRR. Here we connect these topics to psychology's replication crisis (Shrout & Rodgers, 2019), which has recently come to afflict expertise science (Macnamara & Maitra, 2019), by illustrating how differences in data aggregation practices and terminology across studies can foster illusory failures to replicate and stymie building a robust body of empirical evidence linking differences in individuals' attributes to differences in their expertise. To do so we reexamine the data presented in Tables 1 and 2 by combining them in different ways, showing how different decisions using the same data, especially when accompanied by vague linguistic operationalizations, can have serious implications for judging whether prior research has been replicated. Although in the following text we highlight several examples of how conclusions drawn from the same data can vary greatly when those data are recombined in simple but novel ways, all potential aggregations can be found in the Appendix (Tables 1A and 2A).

ES1 featured experts in two domains (politics, wealth acquisition) and each domain featured two "levels" of skill, defined in terms of the rarity and presumed difficulty of achieving the relevant marker. The decision to categorize experts into groups having fortunes of at least $30 million is somewhat arbitrary, though the $1 billion mark is a clear public marker of wealth. The separation of House representatives and senators, however, is based on preexisting, distinct groups that have obvious real world relevance. In both cases, however, more abstract categories could be used to organize the individuals in these groups. Both monetary groups could simply be classified as "wealthy" and both political groups as "members of the U.S. congress," leading to different conclusions about the representation of individuals who attended elite schools versus Harvard. With these new, less well-specified categories the results are: Wealthy individuals (.344 elite, .092 Harvard) and members of Congress (.244 elite, .076 Harvard). For both aggregate criterion groups the numbers are, naturally, closer to the proportions reported for the less select expert groups, given those individuals appear with greater frequency in the samples.

Imagine, now, that Investigators A and B attempt to reproduce the trend toward greater political skill being accompanied by higher cognitive ability. Investigator A gathers a sample of "Congressional members," but it is almost solely comprised of House representatives, while Investigator B's sample contains a larger number of senators than House representatives. Given the nature of Investigator A's sample, the results would closely resemble those of ES1 for House representatives (.206 elite, .066 Harvard)—yet these proportions diverge enough from the .244 (elite) and .076 (Harvard) that Investigator A might conclude that ES1's findings failed to replicate. Investigator B's results would more closely approximate ES1's findings for senators, given their overrepresentation in this hypothetical sample. Once more a conclusion could be that ES1 has failed to replicate. Both Investigators A and B possessed data that *could* have replicated the findings of ES1, had only the nature of the political groups been specified at a less general level than "Congressional members". This is the case even though Investigator A's data suffered from RTRR (due a relative absence of senators) and Investigator B's data were not so restricted. Language imprecision (A. P. Fiske, 1981; D. W. Fiske, 2019) when defining degrees of expertise can have deleterious consequences for replication efforts even when investigators possess data that are capable of overcoming the limits imposed by RTRR; the benefits of proper methodology can easily be undone by ambiguous terminology (Kell, 2018).

More severe difficulties might be encountered if vague terminology was accompanied by different aggregate decisions when trying to replicate the findings of ES2. For example, the trend toward individuals with stronger abilities appearing at a higher rate among holders of advanced degrees would be distorted if the criterion category was changed to the more general "holding an advanced degree", which would include individuals who held a master's *or* doctoral degree. With this alteration, the results would be: Top 1% (.61), top 0.5% (.73), and top 0.01% (.63). If only Cohorts 1 and Cohorts 3 were compared it would look like there is no association between ability and expertise, because the label "holding an advanced degree" does not differentiate between master's and doctoral degrees. If an investigator happened to conduct a study solely using individuals in the top 0.5% and 0.01% it would look like differences in ability are *negatively* associated with differences in skill among experts!

Similar challenges—with attendant adverse consequences for perceived replicability—can occur when predictor groups are ambiguously specified. In ES1 if the less stringent "Elite School" category was expanded to include the more stringent Harvard category the new results for the four expert groups would be: 30^{+}-millionaires (.428), billionaires (.556), House representatives (.272), and senators (.53). In ES2, if the broad categorical label "the top 1%" was chosen without specifying the ranges of abilities in the samples it could lead to highly inconsistent results depending on which ability grouping(s) was present. For instance, in terms of patent-holding, if the label "the top 1%" refers primarily to individuals scoring in the top 1% but below the top 0.5%, the rate of achievement of patent-holding is .054. However, if "the top 1%" also includes a large number of people scoring in the top 0.5% range, the rate of patent-holding is .075, while if "the top 1%" also includes people whose scores place them at or above the .01% level, the patent achievement rate is .087. Finally, if "the top 1%" is primarily comprised of individuals scoring at the top 1% and top 0.01% (but not the top 0.5%) the patent-holding rate is .065. In all cases the group label "top 1%" is accurate—but its imprecision could foster greatly varying findings.

These simple data manipulations illustrate that vague language can inhibit replication even when investigators possess data that produce conclusions consistent with those of prior studies. Macnamara and Maitra (2019) speculated that one of the reasons they failed to replicate Ericsson et al.'s (1993) findings concerning deliberate practice was because the two studies' experts differed in their degrees of skill – implying that the original investigation lacked sufficient detail to allow for the expert samples to be adequately compared. Replicating expertise research findings is challenging enough—due to often being forced to rely on small, range restricted samples—and there is no reason for investigators or prospective replicators to make the task more difficult by describing their samples opaquely.

Conclusion

Studying expertise is difficult. Experts are rare, making it challenging to acquire samples large enough to produce statistically stable results. This challenge is amplified when investigators seek to study differences among experts themselves, necessitating large samples that contain progressively rarer individuals whose skills approach the current limits of human performance. These same difficulties arise when studying individuals whose scores vary within the upper ranges of many human traits, be they cognitive, affective, behavioral, or physical. Regardless of whether researchers take a criterion-centered or predictor-centered approach to studying correlates of differences in expert performance, their task is herculean. We have provided recommendations that we hope will assist investigators in expanding the science of expertise.

Footnote

1. Given the prominence of the nature-nurture debate when studying expertise, when we refer to traits throughout our discussion we use the term solely descriptively ("surface trait"): "Surface traits are not determining tendencies or genotypic mechanisms. Surface traits are descriptions of human tendencies couched in ordinary language" (Wiggins, 1984, p. 182).

Acknowledgement

We thank Samuel McAbee, Thomas Stafford, Brenton Wiernik, and an anonymous reviewer for comments that enhanced the quality of this manuscript.

Author's Declarations

The authors declare that there are no personal or financial conflicts of interest regarding the research in this article.

The authors declare that they conducted the research reported in this article in accordance with the Ethical Principles of the Journal of Expertise.

The authors declare that they are not able to make the dataset publicly available but are able to provide it upon request.

References

Ackerman, P. L. (2014). Facts are stubborn things. *Intelligence, 45*, 104-106.

Arneson, J. J., Sackett, P. R., & Beatty, A. S. (2011). Ability-performance relationships in education and employment settings: Critical tests of the more-is-better and the good-enough hypotheses. *Psychological Science*, *22*, 1336-1342.

Astin, A. W. (1964). Criterion-centered research. *Educational and Psychological Measurement*, *24*, 807-822.

Baker, J., Wattie, N., & Schorer, J. (2015). Defining expertise: A taxonomy for researchers in skill acquisition and expertise. In J. Baker, & D. Farrow (Eds.). *The Routledge handbook of sport expertise* (pp. 145-155). London, UK: Routledge.

Benbow, C. P. (1988). Sex differences in mathematical reasoning ability in intellectually talented preadolescents: Their nature, effects, and possible causes. *Behavioral and Brain Sciences*, *11*, 169-183.

Bogen, J., & Woodward, J. (1988). Saving the phenomena. *The Philosophical Review, 97,* 303-352.

Brooks, D. (2011). *The social animal*. New York, NY: Random House.

Caruso, D. R., Fleming, K., & Spector, E. D. (2014). Emotional intelligence and leadership. In G. R. Goethals, S. T. Allison, R. M. Kramer, & D. M. Messick (Eds.), *Conceptions of leadership* (pp. 93-110). New York, NY: Palgrave Macmillan.

Chi, M. T. H. (2006). Two approaches to the study of experts characteristics. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 21-30). New York, NY: Cambridge University Press.

Chi, M. T. H., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R. J. Sternberg (Eds.), *Advances in the psychology of human intelligence* (pp.1-75). Hillsdale, NJ: Erlbaum.

College Board (2017). *2017: SAT suite of assessments annual report*. Retrieved from https://reports.collegeboard.org/pdf/2017-total-group-sat-suite-assessments-annual-report.pdf

College Board (2018a). SAT results: Class of 2018. *SAT Suite Results: 2018*. Retrieved from https://reports.collegeboard.org/sat-suite-program-results /class-2018-results

College Board (2018b). *SAT: Understanding scores 2018*. Retrieved from https://collegereadiness.collegeboard.org/pdf/understanding-sat-scores.pdf

DeMeyer, T., Bromberg, L., & Hindle, T. (2019, November 17). Jerry Rice, Tom Brady and Lawrence Taylor top our list of the 100 greatest NFL players of all time. *USA TODAY*. Retrieved from https://www.usatoday.com/story/sports/nfl-100/2019/10/01/nfl-100-best-players-all-time/3785514002/

Endsley, M. R. (2006). Expertise and situation awareness. In K. A. Ericsson, N. Charness, P. J. Feltovitch, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 613-652). Cambridge, UK: Cambridge University Press.

Ericsson, K. A. (2006). The influence of experience and deliberate practice on the development of superior expert performance. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 683-704). New York, NY: Cambridge University Press.

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.

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.

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

Fiske, A. P. (2019). The lexical fallacy in emotion research: Mistaking vernacular words for psychological entities. *Psychological Review*.

Fiske, D. W. (Ed.) (1981). *New directions for methodology of social and behavioral science: Problems with language imprecision*. San Francisco, CA: Jossey-Bass.

Frey, M. C., & Detterman, D. K. (2004). Scholastic assessment or *g*? The relationship between the scholastic assessment test and general cognitive ability. *Psychological Science*, *15*, 373-378.

Galton, F. (1869). *Hereditary genius*. London, UK: Macmillan.

Gobet, F., & Campitelli, G. (2007). The role of domain-specific practice, handedness, and starting age in chess. *Developmental Psychology*, *43*, 159-172.

Gobet, F., & Simon, H. A. (1996). Recall of rapidly presented random chess positions is a function of skill. *Psychonomic Bulletin & Review*, *3*, 159-163.

Gould, S. J. (1996). *The mismeasure of man*. New York, NY: Norton & Company.

Gower, B. (1997). *Scientific method: An historical and philosophical introduction*. New York, NY: Routledge.

Grabner, R. H., Stern, E., & Neubauer, A. C. (2007). Individual differences in chess expertise: A psychometric investigation. *Acta Psychologica*, *124*, 398-420.

Gross, M. U. M. (2002). Social and emotional issues for exceptionally intellectually gifted students. In M. Neihart, S. Reis, N. M. Robinson, & S. Moon (Eds.), *The social and emotional development of gifted children: What do we know?* (pp. 19-29). Waco, TX: Prufrock Press.

Haber, R. N., & Haber, L. (2003). Perception and attention during low-altitude high-speed flight. In P. S. Tsang & M. A. Vidulich (Eds.), *Principles and practice of aviation psychology* (pp. 21-68). Mahwah, NJ: Lawrence Erlbaum Associates.

Hoffman, J. R., Tenenbaum, G., Maresh, C. M., & Kraemer, W. J. (1996). Relationship between athletic performance tests and playing time in elite college basketball players. *The Journal of Strength & Conditioning Research*, *10*, 67-71.

Johnson, P. E., Duran, A. S., Hassebrock, F., Moller, J., Prietula, M., Feltovich, P. J., & Swanson, D. B. (1981). Expertise and error in diagnostic reasoning. *Cognitive Science*, *5*, 235-283.

Judkins, T. N., Oleynikov, D., & Stergiou, N. (2009). Objective evaluation of expert and novice performance during robotic surgical training tasks. *Surgical Endoscopy*, *23*, 590-597.

Kell, H. J. (2018). Noncognitive proponents' conflation of "cognitive skills" and "cognition" and its implications. *Personality and Individual Differences*, *134*, 25-32.

Kell, H. J., & Lubinski, D. (2015). Intellectual abilities for counseling interventions, practice, and theory. In P. J. Hartung, M. L. Savickas, & W. B. Walsh (Eds.), *APA handbook of career intervention* (pp. 303-326). Washington, DC: American Psychological Association.

Kell, H. J., Lubinski, D., & Benbow, C. P. (2013). Who rises to the top? Early indicators. *Psychological Science*, *24*, 648-659.

Kozbelt, A. (2007). A quantitative analysis of Beethoven as self-critic: Implications for psychological theories of musical creativity. *Psychology of Music*, *35*, 144-168.

Ledbetter, D. O. (2011, February 26). NFL's physical testing methods scrutinized. *The Atlanta Journal-Constitution*. Retrieved from https://www.ajc.com/sports/football/nfl-physical-testing-methods-scrutinized/q7QLMFGkZ4qbSDaQ6xJ8TP/

le Gall, F., Carling, C., Williams, M., & Reilly, T. (2010). Anthropometric and fitness characteristics of international, professional and amateur male graduate soccer players from an elite youth academy. *Journal of Science and Medicine in Sport*, *13*, 90-95.

Lubinski, D., & Benbow, C. P. (2006). Study of mathematically precocious youth after 35 years: Uncovering antecedents for the development of math-science expertise. *Perspectives on Psychological Science*, *1*, 316-345.

Lubinski, D., Benbow, C. P., & Kell, H. J. (2014). Life paths and accomplishments of mathematically precocious males and females four decades later. *Psychological Science*, *25*, 2217-2232.

Lyons, B. D., Hoffman, B. J., Michel, J. W., & Williams, K. J. (2011). On the predictive efficiency of past performance and physical ability: The case of the National Football League. *Human Performance*, *24*, 158-172.

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.

Macnamara, B. N., & Maitra, M. (2019). The role of deliberate practice in expert performance: Revisiting Ericsson, Krampe & Tesch-Römer (1993). *Royal Society Open Science*, *6*, 190327.

Makel, M. C., Kell, H. J., Lubinski, D., Putallaz, M., & Benbow, C. P. (2016). When lightning strikes twice: Profoundly gifted, profoundly accomplished. *Psychological Science*, *27*, 1004-1018.

McAbee, S. T. (2018). Statistical and methodological considerations for research on expertise: Remedies from Open Science and the Reproducibility Crisis. *Journal of Expertise*, *1*, 136-143.

McAbee, S. T., & Oswald, F. L. (2017). Primer: Statistical methods in the study of expertise. In D. Z. Hambrick, G. Campitelli, & B. Macnamara (Eds.), *The science of expertise: Behavioral, neural, and genetic approaches to complex skill* (pp. 13-30). New York, NY: Routledge.

Meehl, P. E. (2004). Cliometric metatheory III: Peircean consensus, verisimilitude and asymptotic method. *The British Journal for the Philosophy of Science*, *55*, 615-643.

MLB Advanced Media (2019). *Statistics*. Retrieved from http://mlb.mlb.com/stats/sortable.jsp?c_id=mlb#elem=%5Bobject+Object%5D&tab_level=child&click_text=Sortable+Player+hitting&game_type='R'&season=2019&season_type=ALL&league_code='MLB'&sectionType=sp&statType=hitting&page=1&ts=1574056905334&split=&playerType=ALL&timeframe=&active_sw='Y'

Murray, C. (2003). *Human accomplishment: The pursuit of excellence in the arts and sciences, 800 B.C. to 1950.* New York, NY: Harper Collins.

Park, G., Lubinski, D., & Benbow, C. P. (2008). Ability differences among people who have commensurate degrees matter for scientific creativity. *Psychological Science*, *19*, 957-961.

Quetelet, A. (1842). *A treatise on man and the development of his faculties* (R. Knox, Trans.). Edinburgh, Scotland: Chambers.

Rodgers, J. L., & Nicewander, W. A. (1988). Thirteen ways to look at the correlation coefficient. *The American Statistician*, *42*, 59-66.

Shrout, P. E., & Rodgers, J. L. (2018). Psychology, science, and knowledge construction: Broadening perspectives from the replication crisis. *Annual Review of Psychology*, *69*, 487-510.

Simonton, D. K. (1991). Latent-variable models of posthumous reputation: A quest for Galton's G. *Journal of Personality and Social Psychology*, *60*, 607-619.

Simonton, D. K. (1994). *Greatness: Who makes history and why*. New York, NY: Guilford Press.

Stewart III, C., & Reynolds, M. (1990). Television markets and US Senate elections. *Legislative Studies Quarterly*, *15*, 495-523.

Stump, A. (1996). *Cobb: A biography*. Chapel Hill, NC: Algonquin Books.

Teramoto, M., & Cross, C. L. (2018). Importance of team height to winning games in the National Basketball Association. *International Journal of Sports Science & Coaching*, *13*, 559-568.

Thorndike, R. L. (1949). *Personnel selection: Test and measurement techniques*. New York, NY: John Wiley & Sons.

Underwood, B. J. (1975). Individual differences as a crucible in theory construction. *American Psychologist*, *30*, 128-134.

U.S. Census Bureau. (2012a). *Bachelor's degree attainment tops 30 percent for first time, Census Bureau reports*. Retrieved from http://www.census.gov/newsroom/releases/archives/education/cb12-33.html

U.S. Census Bureau. (2012b). *Educational attainment in the United States: 2012 – detailed tables*.

Retrieved from http://www.census.gov/hhes/socdemo/education/data/cps/2012/tables.html
U.S. Patent and Trademark Office. (2011). *U.S. patent statistics chart: Calendar years 1963 – 2011*. Retrieved from http://www.uspto.gov/web/offices/ac/ido/oeip/taf/us_stat.htm
U.S. Patent and Trademark Office. (2015). *General information concerning patents*. Retrieved from https://www.uspto.gov/patents-getting-started/general-information-concerning-patents
Vaci, N., Edelsbrunner, P., Stern, E., Neubauer, A., Bilalić, M., & Grabner, R. H. (2019). The joint influence of intelligence and practice on skill development throughout the life span. *Proceedings of the National Academy of Sciences*, *116*, 18363-18369.
Vaci, N., Gula, B., & Bilalić, M. (2014). Restricting range restricts conclusions. *Frontiers in Psychology*, *5*, 569.
Voss, J. F., Greene, T. R., Post, T. A., & Penner, B. C. (1983). Problem-solving skill in the social sciences. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 17, pp. 165-213). New York, NY: Academic Press.
Wai, J. (2013). Investigating America's elite: Cognitive ability, education, and sex differences. *Intelligence, 41*, 203-211.
Wai, J. (2014a). Experts are born, then made: Combining prospective and retrospective longitudinal data shows that cognitive ability matters. *Intelligence*, *45*, 74-80.
Wai, J. (2014b). What does it mean to be an expert? *Intelligence*, *45*, 122-123.
Wai, J., & Kanaya, T. (2019). Wealth generation as a form of expertise: An examination from 2002-2016 of elite education, cognitive ability, and the gender gap among billionaires. *Journal of Expertise, 2,* 59-76.
Wai, J., & Lincoln, D. (2016). Investigating the right tail of wealth: Education, cognitive ability, giving, network power, gender, ethnicity, leadership, and other characteristics. *Intelligence, 54,* 1-32.
Wai, J. & Rindermann, H. R. (2015). The path and performance of a company leader: An historical examination of the education and cognitive ability of Fortune 500 CEOs. *Intelligence, 53,* 102-107.
Wiggins, J. S. (1984). Cattell's system from the perspective of mainstream personality theory. *Multivariate Behavioral Research*, *19*, 176-190.
World Chess Federation (2019). *Standard top 100 players November 2019*. Retrieved from https://ratings.fide.com/top.phtml?list=men

Received: 1 August 2019
Revision received: 17 December 2019
Accepted: 18 December 2019

Appendix

Table 1A. Summary of Results for Example Study 1, Including Results for Aggregated Categories

Expertise Criterion \ Ability Level	Elite School	Harvard	Elite School or Harvard*
Wealth			
Net worth ≥ $30 million	.338	.090	.428
Net worth ≥ $1 billion	.434	.122	.556
Wealthy Individuals†	.344	.092	.436
Politics			
U.S. House of Representatives Members	.206	.066	.272
U.S. Senators	.410	.120	.530
Members of Congress‡	.244	.076	.320

Note. Percentages for experts in building wealth are based on data from Wai and Lincoln (2016, p. 17, Appendix A). Percentages for political experts are based on data from Wai (2013, p. 206, Table 2).

*Sample size-weighted aggregate of individuals who attended an elite educational institution or Harvard University.
†Sample size-weighted aggregate of individuals with a net worth≥ $30 million and individuals with a net worth ≥ $1 billion.
‡Sample size-weighted aggregate of U.S. House of Representatives members and senators.

Table 2A. Summary of Results for Example Study 2, Including Results for Aggregated Categories

Expertise Criterion Ability Level	Advanced degree[*]	Doctoral degree	Peer-reviewed publication (≥ 1)	NSF grant (≥ 1)	NIH grant (≥ 1)	Patent (≥ 1)
Top 1% (Cohort 1)	.606	.24	.212	.023	.027	.054
Top 0.5% (Cohort 2)	.734	.314	.322	.038	.03	.126
Top 0.01% (Cohort 3)	.634	.44	.24	.06	.03	.15
“Top 1%” (Cohorts 1 – 2)[†]	.645	.337	.204	.023	.023	.063
“Top 1%” (Cohorts 2 – 3)[‡]	.7	.519	.257	.042	.027	.121
“Top 1%” (Cohorts 1, 3)[ɸ]	.611	.278	.19	.027	.024	.065
“Top 1%” (Cohorts 1 – 3)[п]	.643	.351	.209	.028	.024	.075

Note. Percentages for Cohort 1 and 2’s doctoral degrees are based on data from Benbow et al. (2000, p. 475, Table 1) and for all other accomplishments are based on Lubinski et al. (2014, pp. 2218 & 2220, Table 2). Percentages for all Cohort 3’s accomplishments are based on data from Makel et al. (2016, p. 1012, Table 1).

[*]Held a master’s or doctoral degree.
[†]Sample size-weighted aggregate of individuals in the top 1% and 0.5%.
[‡]Sample size-weighted aggregate of individuals in the top 0.5%, and 0.01%.
[ɸ]Sample size-weighted aggregate of individuals in the top 1% and 0.01%.
[п]Sample size-weighted aggregate of individuals in the top 1%, 0.5%, and 0.01%.