Journal of School Psychology 103 (2024) 101269

Contents lists available at ScienceDirect

Journal of School Psychology

journal homepage: www.elsevier.com/locate/jschpsyc

Method considerations for school psychology from longitudinal research on gifted students☆

Jonathan Wai [a,*], Harrison J. Kell [b], Frank C. Worrell [c]

[a] Dept. of Education Reform and Dept. of Psychology, University of Arkansas, Fayetteville, AR, USA
[b] Human Resources Research Organization (HumRRO), Alexandria, VA, USA
[c] Berkeley School of Education, University of California, Berkeley, CA, USA

ARTICLE INFO

Editor: Craig A. Albers
Action Editor: Matthew Reynolds

Keywords:
Gifted education
Longitudinal research
Replication
Spatial reasoning
Predictors
Outcomes

ABSTRACT

This article draws from longitudinal research on gifted students to provide method considerations for school psychology research. First, we provide some background of gifted and talented education in the United States. Then, drawing from multiple longitudinal samples of gifted students, in particular the Study of Mathematically Precocious Youth (SMPY), we illustrate the role of replications, including constructive replications. In the middle two sections, we highlight methodological design features focused first on predictors, and then on outcomes, considering types, magnitude, and breadth. Finally, we provide additional considerations and future directions, including expanding the outcome domain, overcoming the limitations of past gifted and talented research studies, and suggesting possibilities for future research. Our article may help improve school psychology research as well as assist school psychology researchers interested in conducting their own longitudinal studies using gifted samples.

1. Introduction

School psychologists can often be one of the first points of contact for gifted students and their parents and have the technical expertise in measurement and assessment to recognize the individual developmental and educational needs of students who have gifts or talents in specific or more global domains that warrant more specialized and tailored educational opportunities. From that perspective, school psychologists may be an underappreciated part of both the identification and talent development aspects of gifted education. This article introduces some method considerations from longitudinal research on gifted populations for school psychologists and researchers, with the goal of helping school psychologists understand core research and practice findings from the gifted education field and, in addition, providing them with valuable tools and perspectives for studying gifted students themselves.

We do not intend our treatment to be a comprehensive review of longitudinal research designs or the gifted education literature. Instead, we highlight a few important methodological features of longitudinal studies of gifted individuals, with an emphasis on the United States and one longitudinal study known as the Study of Mathematically Precocious Youth (SMPY; Lubinski & Benbow, 2021). We focus on the United States and SMPY as our target audience is school psychology researchers and practitioners working in schools

☆ Special issue for *Journal of School Psychology* (JSP) for a Virtual Special Issue on "Longitudinal Methods to Advance School Psychology Research." https://www.journals.elsevier.com/journal-of-school-psychology/call-for-papers/special-issue-on-longitudinal-methods-to-advance-school-psychology-research.

* Corresponding author.
E-mail addresses: jwai@uark.edu (J. Wai), hkell@humrro.org (H.J. Kell), frankc@berkeley.edu (F.C. Worrell).

https://doi.org/10.1016/j.jsp.2023.101269
Received 18 April 2022; Received in revised form 6 April 2023; Accepted 1 December 2023
Available online 22 December 2023

or educational contexts within the United States. Beyond these contexts, we point readers to studies of gifted students in contexts outside of the United States (e.g., Preckel et al., 2017; Ramos et al., 2023; Wirthwein & Rost, 2011). When appropriate, we direct readers to more advanced and detailed treatments of the issues we explore in this article.

We focus our attention on longitudinal studies because they often allow researchers to draw causal inferences, especially in the context of hypotheses based on robust theories and replicable results. Although experimental studies are considered the gold standard for making causal inferences (Cartwright, 2009), these studies are not always possible in educational research for practical and ethical reasons, and longitudinal designs are a powerful alternative when experiments are not possible. Moreover, findings from longitudinal research can also be illuminating for other reasons, including more fully understanding psychology and human development across the lifespan, with schooling as one part of that broader developmental process. Less considered are the methodological insights that can arise from conducting longitudinal research on unique populations or samples. To that end, in this article we illustrate longitudinal methods that can meaningfully inform school psychology researchers who encounter longitudinal research and who may desire to conduct their own longitudinal research on gifted students. Our focus is largely on SMPY (Lubinski & Benbow, 2000, 2006, 2021) as it is the most contemporary longitudinal study of gifted students to date and is ideal for highlighting methodological considerations for school psychology. However, we include longitudinal research using other databases, also focusing on the gifted, to illustrate patterns of data that replicate.

In the first section of this article, we include a brief background relating to the identification of gifted students in the United States, the SMPY study, and then discuss replications, which are important to all kinds of research. Then, we illustrate replications using SMPY data, discuss replications using other relevant population representative samples of gifted youths, and discuss constructive replication examples. The next two sections highlight methodological design features. The second section focuses on predictors, including their magnitude and breadth. The third section focuses on outcomes, including types, magnitude, and breadth. In the fourth and final section, we note additional considerations and future directions, such as expanding the outcome domain, overcoming limitations in past gifted and talented research studies, and suggest future research before concluding this article.

2. Background

2.1. Identifying gifted students in the US

To help provide background for the reader, we briefly review the context of gifted education in United States' schools and in particular the process used to identify students for the gifted and talented designation. In public schools, gifted identification can take place throughout the K–12 years, but most often occurs in the early grades, and typically, the assessments used and procedures for identification are determined at the district level within the constraints of individual state policies. There is wide variation in state policies (Rinn et al., 2022), meaning there is no uniform method of gifted identification or how that identification process unfolds at the local level. For example, within individual states, there is wide variation in the manner that districts and even individual gifted coordinators and committees can interpret these policies and ultimately implement them in practice. Thus, although cognitive aptitude assessments are often used as part of the identification process and figure prominently in many state policies, they most certainly are not the only aspect of that process (Rinn et al., 2022). Recent treatments have emphasized the importance of testing all students universally as part a screening process (e.g., Card & Giuliano, 2016; Tran et al., 2022).

In addition to the gifted identification process in public schools, there is a nationwide talent search program through which students before age 13 years (i.e., roughly the 7th grade) take the SAT or ACT, which are admissions tests designed for college bound high school juniors or seniors (Assouline & Lupkowski-Shoplik, 2012). Students who obtain high SAT or ACT scores in this talent search process, which are considered an indication of their highly developed cognitive aptitude, are often provided educational opportunities through talent search centers across the United States (e.g., Johns Hopkins University's Center for Talented Youth [CTY], Northwestern University's Center for Talent Development [CTD]). The talent search process is closely aligned with the way the SMPY sample was identified. Finally, there are also many options for gifted K–12 students to be identified and served outside the public school system, including at universities, private schools, special schools for the gifted, or via homeschooling.

2.2. Study of Mathematically Precocious Youth (SMPY)

SMPY includes a sample of over 5000 intellectually talented students, most of whom were purposefully identified prior to age 13 years in the 1970s and followed up many years later (Lubinski & Benbow, 2006, 2021). There are five distinct cohorts in the entire sample. Cohort 1 was selected to be a sample roughly in the top 1% of cognitive reasoning using the SAT as the criterion, Cohort 2 was selected to be a sample roughly in the top 0.5% of cognitive reasoning, and Cohort 3 was selected to be a sample roughly in the top 0.01% of cognitive reasoning. Cohort 4 members scored in the top 3% on any subtest of a grade-level achievement test and Cohort 5 is composed of a sample of top STEM graduate students who were surveyed after their admission to these top graduate programs. In this article, we focus primarily on Cohorts 1, 2, and 3, who were classified as intellectually talented in the early 1970s using the SAT, and on the graduate student cohort (i.e., Cohort 5). Cohorts 1–3 serve as prospective longitudinal samples at different levels of cognitive aptitude, whereas Cohort 5 serves as a retrospective sample for comparison in terms of outcomes. For more details about each of the samples that compose SMPY, see Lubinski and Benbow (2006, 2021).

It is important to note that since its outset, SMPY has been focused on tracking the developmental histories of intellectually precocious individuals and, as such, its findings are most relevant to understanding issues centered on cognitive abilities. As our own review centers on the methodological aspects exemplified by SMPY's design, it too primarily focuses on findings concerning cognitive abilities. This focus should not distract from the fact that intra- and interpersonal attributes (e.g., personality traits; psychosocial factors; social, emotional, & behavioral skills; vocational interests) are also of conceptual and practical importance for all students, both in intellectually gifted and general populations (Heckman et al., 2014; National Research Council, 2012; Soto et al., 2022).[1] Indeed, despite SMPY's primary commitment to the exploration of cognitive abilities and their implications, it too features measures of intra- and interpersonal constructs (e.g., Dauber & Benbow, 1990; McCabe et al., 2020), and we selectively highlight the findings relevant to those constructs throughout our treatment here.

2.3. *The importance of replication*

The importance of replication is acknowledged across many scientific disciplines, although replication has gained much more attention within psychology and education due to what has been termed the "replication crisis" (Makel & Plucker, 2014; Open Science Collaboration, 2015). Perhaps less well known is that methodologists in psychology have been concerned about the lack of attention to replication for quite some time (e.g., Lykken, 1968; Meehl, 1978). The core idea behind replication is simply that a finding should be *repeatable* if it is to be taken as indicative of a genuine advance in knowledge, and that to the extent that findings are generalizable, they should be repeatable in numerous contexts. Findings that do not replicate should result in, at a minimum, a re-examination of assumptions, hypotheses, and methodology and may even raise doubts about the theory guiding the research.

One of the benefits of SMPY is that it was designed to facilitate replication. The only change in Cohorts 1–3 was in selectivity of the sample identified, allowing for an examination of the impact of selectivity while still allowing researchers to examine if findings were replicating. Cohort 4 used a different selection criterion to determine if the criterion measure affected the replication of findings, and Cohort 5 allowed for a retrospective and prospective replication, as participants were identified in graduate school rather than early adolescence and were asked to report on their histories in addition to measuring their post-graduate school trajectories.

2.3.1. *Replication across independent cohorts within SMPY*

As noted above, three of the independent cohorts from SMPY were deliberately selected to be at somewhat different levels of cognitive aptitude: Cohort 1 in the top 1%, Cohort 2 in the top 0.5%, and Cohort 3 in the top 0.01%. Thus, these three cohorts allow us to use replications to ask the following questions: (a) Do higher levels of cognitive aptitude matter for educational attainment, occupational outcomes, and creative accomplishments? and (b) Does the pattern of specific abilities matter for these outcomes (Lubinski & Benbow, 2006, 2021)? In answering the first question specific to educational outcomes, SMPY participants were surveyed 35 years after the study's launch (Lubinski & Benbow, 2006). Among those scoring in the top 1% of math reasoning (i.e., Cohort 1), 88% earned bachelor's degrees, 37% earned master's degrees, and 24% earned doctoral degrees. Within the top 0.5% of math reasoning in SMPY (i.e., Cohort 2), 96% earned bachelor's degrees, 42% earned master's degrees, and 31% earned doctoral degrees. Within the top 0.01% of reasoning (i.e., Cohort 3), 96% earned bachelor's degrees, 39% earned master's degrees, and 51% earned doctoral degrees (see Wai, 2014, for additional summary statistics).

Thus, the findings across multiple samples indicate that higher levels of developed cognitive aptitude are strongly associated with higher educational attainment, even in groups of students at the upper end of the reasoning distribution. In addressing the second part of Question 1 (i.e., occupational outcomes and creative contributions), SMPY shows that as cognitive reasoning level rises, so does level of accomplishment across criteria as diverse as STEM publications, patents, literary publications, and annual incomes. These findings suggest that even within the top 1% of reasoning, more cognitive aptitude matters in terms of long-term educational, occupational, and creative accomplishments (Lubinski, 2009).

One advantage of using the SAT as the measure of developed cognitive aptitude is that each student receives a score on quantitative reasoning and verbal reasoning, which facilitates answering the second question regarding the impact of the pattern of abilities. Park et al. (2007) showed that, within the top 1% of cognitive reasoning (i.e., Cohort 1), individuals with stronger verbal than mathematical reasoning were distinguished by long-term trajectories more aligned with the humanities relative to STEM, earning terminal 4-year and master's degrees, doctoral degrees, holding tenure-track positions, and producing books in humanities fields. The pattern was reversed for individuals whose profiles favored quantitative reasoning, who made more contributions in STEM fields relative to the humanities.

Kell et al. (2013) replicated this pattern of findings specifically in Cohort 3. Their findings mirrored those of Park et al. (2007) and were found to extend to individuals' occupations, peer-reviewed publications, and other creative accomplishments, such as creating software programs and publishing poems. Noteworthy is the fact that SMPY's Cohort 3 is comprised of participants solely in the top 0.01% of cognitive reasoning, meaning that 94% of individuals' weaker cognitive aptitudes were still in the top 1% of reasoning, placing them above most PhD holders in any discipline (Wai et al., 2009). Thus, despite the fact that this sample possessed cognitive aptitude that would likely have allowed them to be successful in any field, most Cohort 3 participants chose to pursue educational, occupational, and creative pathways aligned with their reasoning strength (i.e., quantitative vs. verbal). So, a second major replicable finding among intellectually gifted participants is that individuals with differing intellectual profiles tend to occupy different educational, occupational, and creative niches.

[1] U.S. researchers in education policy have been influenced by education economists in referring to these attributes as "noncognitive", but this terminology is both widely lamented (Messick, 1979) and misleading (Kell, 2018). We thank an anonymous reviewer for pointing out that the label "noncognitive" can be unhelpful because it may implicitly center discussion of human attributes around cognitive aptitudes, perhaps suggesting to some that attributes other than cognitive aptitudes are intrinsically less valuable by relegating them to a single, undifferentiated category.

Bernstein et al. (2019) extended this line of research by focusing on contributions related to being leaders in professional fields, such as winning major awards, earning substantial amounts of grant funding, or being a professor at a research-intensive university. Once more, the pattern replicated: Within the top 1% of reasoning, leaders in humanities fields tended to have stronger verbal than mathematical reasoning, whereas leaders in STEM fields tended to have stronger mathematical than verbal reasoning. In summary, the initial study and the multiple replications indicate that both *level* and *pattern* of cognitive abilities contribute to achievements in educational, occupational, and creative domains.

2.3.2. Replication across independent gifted samples

Additional studies addressing these questions have been conducted with independent cohorts with similar reasoning capacity. The following two examples illustrate replications of SMPY findings with data from (a) an equally select (top 0.01%) but independent sample of gifted youths from the Duke Talent Identification Program (TIP) and (b) the top 1% of a population-representative, random sample of students from the United States. The first replication we summarize is with an independent Talent Search sample at TIP, also selected using the SAT taken before age 13 years (see Makel et al., 2016). The TIP sample was surveyed again at age 40 years. Broadly, the patterns across TIP replicated those across SMPY: 37% earned doctorates, 7.5% had academic tenure (with 4.3% at Research 1 universities), 9% had obtained a patent, and many were leaders of their organizations. Moreover, the structure of creative accomplishments as a function of math and verbal specific reasoning patterns was also replicated across these select samples, with TIP participants also pursuing life trajectories aligned with their reasoning strength. Thus, the pattern of correlations across highly select samples of gifted students between early talent and adult accomplishments is quite similar.

One can argue that both the SMPY and TIP samples, identified through the Talent Search, are unique. Would findings from SMPY and TIP, which are non-random samples deliberately selected to focus on a portion of the intellectually talented population largely identified as middle schoolers in the 1970s, replicate with a random sample of intellectually talented students? Wai (2014) examined this question by drawing from the top 1% in cognitive scores of Project Talent, which was a stratified random sample of the nation's high school students first tested in Grades 9–12 in the 1960s and followed up after their graduation. Table 1 in Wai (2014) used data from SMPY and Project Talent to show that the percentages within the top 1% of reasoning in both samples were remarkably similar in earned bachelors, masters, and doctorate degrees, showing replication across multiple independent cohorts. Table 2 in Wai (2014) shows that even within the top 1%, higher cognitive reasoning was related to a higher rate of earning doctorates in both the SMPY and Project Talent samples. Overall, then, these studies show replication for higher education degree attainment across a nonrandom and random sample of the top 1% in cognitive reasoning.

2.3.3. Constructive replication examples

The purpose of constructive replication, according to Lykken (1968), is to vary different construct-irrelevant design features across studies while keeping the focal constructs of interest the same across each study. For example, Wai et al. (2009) illustrated constructive replications of the relationship between spatial reasoning and numerous educational and occupational outcomes by showing how these patterns of associations found in prior research (Super & Bachrach, 1957) replicated in the SMPY (Shea et al., 2001) and the Project Talent samples. Even though spatial reasoning was measured in different ways using distinct measures at different time points, the samples were of different ages and selectivity levels at the point of initial testing, outcomes varied from high school course preference to occupational attainment, and investigators were independent groups conducting studies for different purposes, the correlational structure between spatial reasoning and numerous educational and occupational outcomes was quite similar.

Construct irrelevant design features might constitute variation across the different datasets collected for very different purposes on everything from initial testing to follow up information collected at later time points. Project Talent, for example, included numerous measures of abilities that were not focused on spatial reasoning and it was a stratified random sample of the United States. SMPY, in contrast, focused on the use of the SAT in addition to one type of spatial reasoning measure for initial talent identification, and the sample was not random (Shea et al., 2001). Super and Bachrach (1957) was a research synthesis of multiple types of studies that included spatial reasoning measures but also had design features that were completely irrelevant to spatial reasoning as a construct, and, despite this, these researchers found patterns between spatial reasoning and outcomes similar to those reported for the SMPY and Project Talent samples collected later in time. Thus, the Wai et al. (2009) study illustrated constructive replications across the last 50 plus years for spatial reasoning.

More recent research has taken a similar approach to constructive replications across a 60-year period by leveraging population representative, longitudinal studies that also included measures of spatial reasoning (Lakin & Wai, 2020; Wai & Lakin, 2020). These studies drew from Unietd States' samples, specifically Project Talent (first cohort identified in 1960), High School and Beyond (1980), and the National Longitudinal Study of Youth (1997). In the first study, Lakin and Wai (2020) examined whether the correlational patterns across different measures of spatial reasoning and educational and social outcomes (across these three independent samples) would replicate. Overall, students with spatial reasoning strengths, relative to those with mathematical or verbal strengths, tended to have more academic challenges, such as difficulty reading, poorer study habits, and other behavioral issues. Additionally, spatially talented students tended to complete college degrees at a lower rate than their mathematically and verbally talented counterparts (Lakin & Wai, 2020).

In the second study, Wai and Lakin (2020) used the same samples to understand whether the proportions of different populations of gifted students from various demographic groups would be uncovered to a similar extent across various cognitive and noncognitive/psychosocial selection tools. Overall, although there was variation, findings were replicated for spatial reasoning (Wai & Lakin, 2020). Because these replications were on multiple independent samples across time using different measures and outcomes but focused on the same focal constructs, the replications are particularly robust (Lykken, 1968).

3. Methodological design features: predictors vs. outcomes

We further divide longitudinal studies into two categories (cf., Kell & Wai, 2019; Sackett et al., 2012).[2] In *predictor-focused* longitudinal studies, focal constructs are identified at the outset and their associations with other variables are examined over the timeframe of the study, with later time points often showcasing important outcomes the original constructs are predictive of, especially due to some causal pathway. However, even when other variables in the study are considered important or interesting in their own right, the focus of the study remains on the constructs identified at its outset with the degree of interest in other variables examined being driven by their association or lack thereof with the focal constructs. For example, SMPY features hundreds of important variables, but they are of interest in the context of SMPY only because of their relationship with SMPY's focal constructs (i.e., cognitive reasoning abilities), which is what drove the initial selection of the participants. Thus, in predictor-focused studies, developing the nomological network of focal constructs examined at the outset of the study is the central concern.

In *outcome-focused* longitudinal studies, the central concern is developing a better understanding of the nomological network of variables that occur later in a developmental sequence and typically acknowledged to be the result of a complex set of causal determinants. For example, the Grant Study of Harvard undergraduates—and its matching study of individuals of lower socioeconomic status—was launched to uncover possible causal determinants of healthy aging (Vaillant, 2000). In this case, variables indicative of healthy aging (e.g., ailments, life expectancy) were the ultimate variables of interest, with other variables in the study—including those measured initially—being of interest primarily because of their potential capacity to ultimately understand and forecast healthy aging.

Outcome-oriented research is intended to identify variables that aid in understanding and forecasting outcome variables as comprehensively as possible, whereas predictor-oriented research is concerned with understanding the focal constructs of interest identified at the outset of the study. If, for example, a longitudinal study of cognitive reasoning (i.e., predictor-focused) indicated that cognitive abilities were poor predictors of career satisfaction, the investigators would likely not add measures of variables that might better predict career satisfaction in their next wave, as career satisfaction, although an important variable, is not the key variable of interest in the study. However, if the investigators of a longitudinal study of career satisfaction (i.e., outcome-focused) found that their current battery of measures was not doing a good job accounting for variance in career satisfaction as the study progressed, they would likely add theoretically- and empirically-informed variables during future waves (and when planning future longitudinal research) in the hope of better understanding and predicting career satisfaction itself.

We introduce these distinctions to clarify the terminology we use throughout the following sections as we regularly refer to some variables—in the case of SMPY, typically cognitive aptitudes—as predictors and others as outcomes, especially when making reference to methodological choices concerning different sets of variables in longitudinal research. Moreover, we also believe more broadly that carefully thinking about the distinctions between predictors and outcomes is useful in helping investigators in school psychology plan their own future longitudinal investigations. Distinguishing explicitly between predictors and outcomes is also key to recommendations that we highlight in the following sections.[3]

3.1. Magnitude of predictors

In longitudinal research, for an accurate association to be present between focal constructs and other variables in the study, the variability in the focal constructs in the study must match that in the population of interest the longitudinal sample is drawn from. When the measurement of variability of a construct is limited relative to its variability in the population of interest, there is range restriction in that construct, which can often reduce the observed association between that variable and others (Kell & Wai, 2019; Sackett & Yang, 2000). This concern is of particular importance in intellectually gifted samples, especially those like SMPY, because a full one-third of the cognitive reasoning range exists in the top 1% of the distribution (Lubinski & Benbow, 2000). Consequently, if assessment procedures are not sufficiently difficult, they will fail to register differences among individuals in the top 1% of cognitive aptitude—that is, they will suffer from a ceiling effect, such that variation in the construct of interest at the high end is not associated with differences in scores at the high end of difficulty of the instrument (Wang et al., 2008). A ceiling effect results in range restriction on the predictor, artificially reducing its observed association with other variables.

As noted, awareness of ceiling effects and range restriction is critical when studying gifted populations because there are differences in achievement even within the top 1% of reasoning. In the case of SMPY, these differences would have been masked if the study's assessments were not able to capture variability in cognitive reasoning at the high end. As noted previously, 24% of Cohort 1, 31% of Cohort 2, and 51% of Cohort 3 earned doctoral degrees—differences in attainment that would have been masked had the initial reasoning assessments been unable to differentiate individuals *within* the top 1% of reasoning. Similarly, Cohort 2 participants obtained NSF grants at over 1.5 times the rate of Cohort 1 participants and Cohort 3 participants obtained NSF grants at over 2.5 times the rate of Cohort 1 participants—differences that, again, would have been obscured had SMPY's assessment lacked sufficient measurement precision (Kell & Wai, 2019).

One solution to avoid range restriction on predictors when working with extremely high scoring samples is to make the test very

[2] We believe these two designations are practically useful for designing longitudinal studies and categorizing preexisting research but do not claim that they exhaustively define all longitudinal studies nor that all such studies will neatly fit into either category.

[3] We define a predictor as a property of persons or their immediate environments that has relevance for forecasting or understanding individuals' future behavior or properties (cf., Klimoski, 1993, p. 101).

difficult. The talent search model is one way to increase the headroom of the test. In talent searches across the United States, at locations such as the Johns Hopkins University CTY and the Northwestern University CTD, students take the SAT or ACT designed for high school juniors or seniors in the seventh grade (Assouline & Lupkowski-Shoplik, 2012). This approach avoids range restriction on the measure and allows students who might typically score at the ceiling on an on-grade standardized test to display their full potential.

Many current testing systems in schools that focus on academic growth also can be used for the purpose of ensuring there is sufficient headroom on the measure. For example, many norms on tests such as Measures of Academic Progress (MAP) designed by NWEA provide data on whether, for example, a Kindergartner is reading at a much higher level based on computer adaptive assessments. The talent search data illustrate that there is likely range restriction on the SAT and ACT as taken by extremely academically capable 11th and 12th graders (Lubinski & Benbow, 2000; Wai, 2012). The core takeaway is that when working with intellectually talented samples, ensuring that there is sufficient headroom on the test is key if one wishes to use that test as a predictor of variables to be studied in the future stages of longitudinal research.

3.2. Breadth of predictors

SMPY originally focused on mathematical and verbal reasoning due to the use of the SAT as the measurement tool for identification. However, SMPY eventually expanded the abilities it assessed, drawing on Guttman's radex model (Lubinski, 2000; Lubinski & Benbow, 2000). The radex model includes the primary specific aptitudes of mathematical, verbal, and spatial reasoning as core to the hierarchical model of cognitive abilities (Carroll, 1993).

For Cohort 2 participants, SMPY investigators added measures of visuospatial reasoning, an aptitude that has been underappreciated for many decades (Kell & Lubinski, 2013), despite evidence for its psychological importance having been apparent since at least the late 19th century (e.g., Binet, 1892). Although the positive manifold dictates that the primary cognitive aptitudes will typically be positively correlated (Borg, 2018), the mean intercorrelation is roughly 0.30 (Carroll, 1993)—and lower in gifted samples (Reynolds, 2013)—meaning that individuals can vary widely across specific abilities. Within Project Talent, for example, about 70% of the top 1% of scorers in spatial reasoning are not in the top 1% of math or verbal reasoning based on a population representative sample (Wai et al., 2009; Webb et al., 2007), suggesting particular utility in measuring spatial reasoning in addition to more typically assessed aptitudes.

SMPY's studies investigating the influence of spatial reasoning on long-term outcomes have consistently affirmed the value of examining cognitive reasoning capacities beyond verbal and quantitative. Shea et al. (2001) found that participants whose intellectual profiles featured stronger spatial and quantitative than verbal reasoning tended to earn undergraduate and graduate degrees in STEM disciplines—in addition to finding employment in such disciplines—whereas individuals evincing the opposite intellectual profile tended to gravitate toward educational and occupational opportunities in the arts and humanities. Kell et al. (2013) extended this work to creative outcomes, including the production of patents and peer-reviewed scholarly publications. Results mirrored those of Shea et al., with SMPY participants with stronger verbal than spatial and quantitative reasoning tending to publish in the arts and humanities and those with stronger visuospatial than verbal reasoning tending to publish in STEM disciplines or produce patents. They further showed that inclusion of spatial reasoning beyond verbal and mathematical reasoning accounted for an additional 7.6% of the variance in predicting creative outcomes more than 30 years after participants took the assessments.

It is important to note that, consistent with recommendations to measure cognitive reasoning constructs using multiple types of content (Lakin & Kell, 2020), Shea et al. (2001) assessed spatial reasoning using two indicators (i.e., mechanical reasoning and spatial relations), whereas Wai et al. (2009) assessed it using four indicators (i.e., three-dimensional spatial visualization, two-dimensional spatial visualization, mechanical reasoning, and abstract reasoning). To assess associations with constructs that are presumed to range from somewhat to highly general in nature, not only should there be breadth in predictors, but breadth in the content of how these predictors are assessed, as well.

Although SMPY's focal constructs are cognitive aptitudes, its participants have been comprehensively assessed in ways consistent with Cattell's (1957) recommendations to gather L-data (i.e., life record), T-data (i.e., tests and other objective assessments), and Q-data (i.e., questionnaires).[4] The initial selection of SMPY participants was based on T-data and the detailed information about many aspects of their lives that has been gathered at each timepoint (e.g., income, number of children, occupational title, relationship satisfaction) constitute L-data (e.g., Kell et al., 2022; Lubinski et al., 2006, 2014). SMPY's Q-data include participants' responses to surveys tapping a wide range of psychosocial constructs, including personality traits, vocational interests, work preferences, and life values (Lubinski et al., 2014; McCabe et al., 2020). Inclusion of such a wide array of assessments not only helps provide a comprehensive portrait of participants' lives over time, it aids in forecasting and understanding important outcomes examined later in the course of the study. For example, McCabe et al. (2020) showed that leaders in STEM fields, compared to non-leaders, differed in their theoretical values, investigative interests, dominance, and self-confidence.

Regarding expanding the breadth of predictors based on population representative longitudinal samples outside of SMPY, the Wai and Lakin (2020) article reviewed earlier is relevant. These researchers examined how the breadth of predictors across cognitive and non-cognitive domains might be expanded to include measures beyond the typical mathematical and verbal reasoning aspects found on most standardized tests. They found that in addition to including spatial reasoning, measures of creativity, the personality trait of conscientiousness, and artistic and leadership interests (as measured in the Project Talent sample they used) might improve the

[4] Block and Block (2006) further recommended the inclusion of observer reports (O-data) in longitudinal research.

representation of historically underserved and low-income students. Spatial reasoning and conscientiousness are well studied constructs and perhaps more likely to replicate, but the other findings were tentative. As Snow (1999) essentially argued, to the extent that a wider range of aptitudes become represented in education, they may then become recognized as goals of education. Of course, if a wider variety of attributes beyond mathematical and verbal aptitudes (e.g., foreign language aptitude, ideational fluency, inductive reasoning) is selected to be explicitly assessed in educational contexts, the selection should be rooted in well-established structures of cognitive, interpersonal, and intrapersonal constructs (e.g., Cattell-Horn-Carroll model of cognitive abilities; five factor model of personality traits; Keith & Reynolds, 2010; Widiger, 2017).

4. Methodological design features: outcomes

Ideally, investigators should put as much care into choosing the outcomes of their longitudinal studies and deciding how to measure those outcomes as they do when selecting the predictors and other focal constructs that they study at the beginning stages of research. To the extent that a longitudinal study is concerned with identifying indicators of “success” (e.g., academic, career, interpersonal, mental health, physical health), investigators must think carefully first about how they want to define success conceptually, then subsequently consider how they want to assess success in the various manifestations they choose (Kell, 2022). In the following sections, we describe how these choices can affect the observed associations between outcomes and the constructs measured at the earlier timepoints of a longitudinal study.

4.1. Types of outcomes

SMPY features a sequence for classifying outcomes first articulated by R. L. Thorndike (1949) who specified three types of outcomes defined both in terms of the timespan in which they become available and their comprehensiveness in representing notions of success. This sequence of outcomes was originally formulated under the auspices of determining how successful personnel selection procedures were, but they are highly applicable to longitudinal research, both in school psychology and beyond. The most important outcomes are *ultimate outcomes*, which are comprehensive, and often multidimensional, variables that represent rationally defined end states for the study in question. These are, fundamentally, conceptual variables that investigators choose, or construct, based on the state of the field and the nature of the research they are conducting. These conceptual outcomes are then operationalized in terms of measurable variables. In SMPY, information about a variety of ultimate outcomes was gathered when participants were in their 50s, a time when most individuals had completed their educations, were well into their professional careers, and were in their peak years of earnings and productivity (Lam et al., 2012; Lubinski et al., 2014). Examples of ultimate outcomes gathered at this time include highest degree obtained, income, occupational status (e.g., Fortune 500 CEO, tenure at a research-intensive university), and number and type of creative accomplishments (e.g., books authored, patents, peer-reviewed publications).

Ultimate outcomes are very useful both conceptually and practically, but they are often difficult to gather information about and it takes a substantial amount of time for those data to become available as they take many years to manifest. For example, many school psychology researchers are probably interested in the highest level of educational attainment their participants reach or the mental and physical health of their participants after they have completed their educations. However, in many cases, it will not be feasible to gather information about those outcomes at all or investigators will want details about their participants' standings on outcomes before these distal markers are finally reached. For example, school psychologists may be interested in whether a classroom intervention or new type of educational practice results in improvements in the earnings of their participants at midlife but are unable to wait for such details to become available before carrying out any type of evaluation of their efforts (see Watts et al., 2019, for discussion on incentivizing this kind of long-term evaluation).

Thus, longitudinal researchers also typically use intermediate outcomes. *Intermediate outcomes* are variables that are judged to be related to ultimate outcomes conceptually or empirically, but they become available before ultimate outcomes are reached while also being separated from the beginning of the investigation by a substantial amount of time. Intermediate outcomes may be related to ultimate outcomes on theoretical, logical, or statistical grounds. For example, if an investigator wants to study how well various student characteristics predict success among attorneys, they could initially examine how well these characteristics predict passing the Bar Exam as doing so is a prerequisite for becoming a practicing attorney. SMPY features a wide variety of outcomes that can be considered intermediate, such as the occupational titles, incomes, and highest degree obtained by participants when they were in their early 30s (Benbow et al., 2000; Lubinski et al., 2006).

Finally, there are *immediate outcomes*, which are variables that are presumed to be related to intermediate and ultimate outcomes, but about which data are available at the beginning of the longitudinal investigation or relatively soon after it begins. In school psychology research focused on long-term academic success (e.g., highest degree obtained) of current high school students, immediate outcomes might include grades in individual classes, cumulative GPA at the end of each academic year or upon completion of high school, graduating after 4 years of high school attendance, status of the undergraduate institution students attended, and highest degree pursued after finishing high school (e.g., associate's vs. bachelor's vs. no degree). Much of education research uses immediate outcomes such as test scores or test score growth, for example.

SMPY contains an exemplary case of the sequence from intermediate to ultimate criteria. Achter et al. (1999) studied participants' college majors and whether they could generally be classified as falling into the arts/humanities, STEM, or other area in their early 20s; Wai et al. (2005) examined whether participants' occupations fell into these same three groupings in their early 30s; and, finally, Bernstein et al. (2019) examined the fields in which participants attained eminence in their careers in their 50s. The first two can be considered examples of intermediate outcomes, whereas the final study concerns outcomes that many would consider to be ultimate outcomes.

4.2. Magnitude of outcomes

The same amount of careful thought that goes into developing methods that adequately reflect the degree of individual differences in predictor variables ideally should go into developing measures of outcomes. Artificially restricted variation in outcome variables due to poor measurement will reduce the observed association between predictors and those outcomes even if the predictors are properly assessed. Just as variation in scores on predictors can be truncated if assessments lack appropriate headroom, so too can scores on outcomes.

A variety of strategies can be employed to ensure that adequate variability in outcomes can be observed (Kell & Wai, 2019; Park et al., 2007). One approach is to employ continuous rather than categorical measures, which allow more precision. When categorical outcomes are used, developing fine-grained categories that can be rank-ordered in terms of difficulty or rarity of attainment can reveal associations with predictors that would otherwise remain hidden. For example, depending upon when the longitudinal study is initiated, a viable immediate outcome in school psychology research might be earned acceptance to an undergraduate institution; however, the precision of this outcome could be expanded by indexing the number of acceptances, the number of acceptances to highly ranked schools (e.g., top 25, top 10), the number of academic scholarships received, and the amount of funding provided by those academic scholarships. In some studies, "stretching out" outcomes in this way may reveal little appreciable variation, but in others it could reveal important differences.

As indicated in the immediate–intermediate–ultimate sequence, time plays an important role in conceptualizing and measuring outcomes. Thus, focusing on outcomes that take a substantial amount of time to emerge is a second strategy for ensuring adequate *headroom* on such variables. In SMPY, 88% of Cohort 1 participants, 96% of Cohort 2 participants, and 96% of Cohort 3 participants obtained 4-year degrees (Lubinski & Benbow, 2006). If investigators had stopped there, the apparent relationship between differences in cognitive reasoning in the top 1% and differences in educational attainment would have been regarded as relatively trivial, and, indeed, it would have appeared that cognitive reasoning beyond the top 0.05% was not important for educational achievement. As the study allowed adequate time for the participants to complete their educational journeys, however, differences in attainment emerged: 24% of Cohort 1 participants, 31% of Cohort 2 participants, and 51% of Cohort 3 participants earned doctoral degrees. Sufficient headroom to capture substantive variability in educational attainment was created by allowing enough time to pass for differences in accomplishment to accumulate. The study also used an outcome with enough precision to capture those differences when they did arise. Finally, the outcome was a sufficiently difficult to achieve—and, thus, rare—criterion that would be relatively unlikely to be reached by the majority of the general population (i.e., the doctoral degree attainment rate in the United States' general population is approximately 2%; U.S. Census Bureau, 2012a, 2012b).

4.3. Breadth of outcomes

A broad sampling of assessments is needed to fully capture the psychological attributes of human beings, both in the general population and among the intellectually gifted. For the long-term implications of those attributes to be revealed, however, an equally broad sample of outcome variables must be assembled. For example, we noted the consistent tendency for individuals with stronger verbal reasoning than quantitative and visuospatial reasoning to gravitate toward the humanities and those with the opposite profiles to gravitate toward STEM (Kell et al., 2013; Wai et al., 2009). The power of these specific aptitudes to forecast the *types* of careers and achievements that SMPY participants have would have gone undiscovered if the *fields* that individuals earned their degrees in and pursued their employment had not been categorized.

Moreover, the many varied ways in which developed verbal, mathematical, and spatial reasoning manifest in people's lives would not have been clear without the fine-grained outcome variables featured in these investigations, including not only detailed categorizations of disciplines within the broad categories of the humanities and STEM (e.g., atmospheric sciences, folklore, petroleum engineering, theology), but also the diverse creative outputs that participants produced within these fields (e.g., patents for Fortune 500 companies, sculptures, software programs, theater productions; Kell et al., 2013; Makel et al., 2016; Wai et al., 2005).

5. Additional considerations and future directions

In the preceding sections, we used the design and results of SMPY to outline beneficial practices for conducting longitudinal research that may be helpful in improving the practice of such research in school psychology. In the following sections, we briefly explore some additional issues school psychologists may want to consider when conducting longitudinal research.

5.1. Expanding the outcome domain

In our recommendations drawn from SMPY, we noted the importance of broadly sampling outcome data when conducting longitudinal research. Over its long history, SMPY has largely focused on the performance of its participants in terms of their achievements in the educational and occupational spheres, with the subsequent creative accomplishments that flow from them. Currently, however, work is underway to expand the outcome space for SMPY even beyond these achievements to include measurements of participation in civic life and contributions to the broader social good (e.g., volunteering activities; McCabe et al., 2018) in recognition of the many varied ways that human beings can contribute to the world around them. Broadly, lifespan developmental research on gifted populations like SMPY, but also population representative samples, are important because there are such a wide range of ways in which talented students can develop their expertise (Subotnik et al., 2011). What SMPY shows in particular is that even students deliberately selected on reasoning measures early on tend to have quite a wide range of eventual occupational outcomes not isolated to any one domain; thus, the breadth of outcomes is incredibly important to capture innovation of all kinds.

Taking a similar approach in school psychology research may be useful for more fully identifying the scope of influence of variables assessed at the outset of longitudinal studies, especially considering rising appreciation of the importance of psychosocial or socio-emotional skills in educational environments (Eklund et al., 2018; Olszewski-Kubilius et al., 2015). On the predictor side, variables that may appear to provide little value in forecasting academic outcomes may be highly valuable in understanding outcomes such as relations with classmates and teachers, mental and physical health, and contributions to the broader school environment (e.g., participation in extracurricular activities). On the outcome side, expanding the scope of variables investigators are interested in can lead to a broader understanding of students' lives and help put into perspective what the fundamental goals of educational systems ultimately are.

5.2. Overcoming the limitations of past gifted studies

Especially in the context of the United States, SMPY has provided numerous insights about gifted students, talent identification, and talent development (Lubinski & Benbow, 2021). However, despite many strengths, it is a study conducted at a particular time segment in history and with students largely drawn from a particular talent identification procedure in the United States. Thus, the sample is not representative of racially, ethnically, and linguistically minoritized individuals, as at the time the sample was identified, many of these individuals had not had the opportunity to fully develop their cognitive aptitudes to be considered gifted or talented (Hair et al., 2015). Issues of structural inequalities and equity largely impacting minoritized students (Peters, 2021) are also aspects that remain unaddressed by the studies discussed here.

Additionally, some critics have noted that similar to Terman's (1925) longitudinal study of giftedness, participants were deliberately selected for and knew about their abilities and may have been provided additional supports (Ericsson, 2014). However, population representative samples of giftedness that were randomly sampled show quite similar proportions of higher degrees earned (e.g., bachelors, masters, PhDs) as a function of reasoning level, suggesting that knowing about one's abilities or supports may not be as important as assumed.

Another possible limitation is the SAT as a measure of developed cognitive reasoning. The purpose of using the SAT at an earlier age (age 13 years compared to 17 or 18 years) for talented students was done with the rationale that the measure might function as more of a reasoning test for talented younger students (Benbow, 1988). Given that the SAT (Frey & Detterman, 2004) and ACT (Koenig et al., 2008) have been shown to largely measure general cognitive aptitude, this seems reasonable, but this approach may still be limited in various aspects.

5.3. Future research and conclusions

School psychologists conduct many studies that focus on immediate rather than intermediate or ultimate goals, but there are many interventions that would benefit from longitudinal research. Future research like SMPY, but extended to larger and more culturally and linguistically diverse samples not only in the United States but also around the world, might be quite fruitful. Multidisciplinary approaches from various fields and methodological tools for studying giftedness and talent development (Wai & Guilbault, 2022) would also help improve the research questions asked and data sources used to address gifted education research but also perhaps how gifted education research can be applicable to other fields and in practice such as for school psychology. If there was a way to help bring together more school psychology and gifted education researchers and practitioners to learn from one another and better understand how each approach might benefit the other, this collaboration would enhance the quality and applicability of research on all students in schools. This especially may help inform how school psychologists can play a larger and even more impactful role in helping identify and support gifted students, especially those from low-income, minoritized, and linguistically diverse backgrounds who have not had the full opportunity to develop their talents. Drawing from longitudinal design innovations from gifted education research may thus be helpful to school psychology more broadly and school psychologists who work with gifted students in particular, and we hope that these two disciplines may form fruitful collaborations when considering aspects introduced in this article.

References

Achter, J. A., Lubinski, D., Benbow, C. P., & Eftekhari-Sanjani, H. (1999). Assessing vocational preferences among gifted adolescents adds incremental validity to abilities: A discriminant analysis of educational outcomes over a 10-year interval. *Journal of Education & Psychology, 91*(4), 777–786. https://doi.org/10.1037/0022-0663.91.4.777.

Assouline, S. G., & Lupkowski-Shoplik, A. (2012). The talent search model of gifted identification. *Journal of Psychoeducational Assessment, 30*(1), 45–59. https://doi.org/10.1177/0734282911433946.

Benbow, C. P. (1988). Sex differences in mathematical reasoning ability in intellectually talented preadolescents: Their nature, effects, and possible causes. *The Behavioral and Brain Sciences, 11*(2), 169–232. https://doi.org/10.1017/S0140525X00049244.

Benbow, C. P., Lubinski, D., Shea, D. L., & Eftekhari-Sanjani, H. (2000). Sex differences in mathematical reasoning ability at age 13: Their status 20 years later. *Psychological Science, 11*(6), 474–480. https://doi.org/10.1111/1467-9280.00291.

Bernstein, B. O., Lubinski, D., & Benbow, C. P. (2019). Psychological constellations assessed at age 13 predict distinct forms of eminence 35 years later. *Psychological Science, 30*(3), 444–454. https://doi.org/10.1177/0956797618822524.

Binet, A. (1892). Mental imagery. *Fortnightly, 52*(307), 95–104.

Block, J., & Block, J. H. (2006). Venturing a 30-year longitudinal study. *The American Psychologist, 61*(4), 315–327. https://doi.org/10.1037/0003-066X.61.4.315.

Borg, I. (2018). A note on the positive manifold hypothesis. *Personality and Individual Differences, 134*, 13–15. https://doi.org/10.1016/j.paid.2018.05.041.

Card, D., & Giuliano, L. (2016). Universal screening increases the representation of low-income and minority students in gifted education. *Proceedings of the National Academy of Sciences, 113*(48), 13678–13683. https://doi.org/10.1073/pnas.1605043113.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press.

Cartwright, N. D. (2009). What is this thing called 'efficacy'? In C. Mantzavinos (Ed.), *Philosophy of the social sciences: Philosophical thought and scientific practice* (pp. 185–206). Cambridge University Press.

Cattell, R. B. (1957). *Personality and motivation structure and measurement*. World Book Company.

Dauber, S. L., & Benbow, C. P. (1990). Aspects of personality and peer relations of extremely talented adolescents. *The Gifted Child Quarterly, 34*(1), 10–14. https://doi.org/10.1177/001698629003400103.

Eklund, K., Kilpatrick, K. D., Kilgus, S. P., & Haider, A. (2018). A systematic review of state-level social–emotional learning standards: Implications for practice and research. *School Psychology Review, 47*(3), 316–326. https://doi.org/10.17105/spr-2017.0116.v47-3.

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. https://doi.org/10.1016/j.intell.2013.12.001.

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*(6), 373–378. https://doi.org/10.1111/j.0956-7976.2004.00687.x.

Hair, N. L., Hanson, J. L., Wolfe, B. L., & Pollak, S. D. (2015). Association of child poverty, brain development, and academic achievement. *JAMA Pediatrics, 169*(9), 822–829. https://doi.org/10.1001/jamapediatrics.2015.1475.

Heckman, J. J., Humphries, J. E., & Kautz, T. (2014). *The myth of achievement tests: The GED and the role of character in American life*. University of Chicago Press.

Keith, T. Z., & Reynolds, M. R. (2010). Cattell–horn–Carroll abilities and cognitive tests: What we've learned from 20 years of research. *Psychology in the Schools, 47*(7), 635–650. https://doi.org/10.1002/pits.20496.

Kell, H. J. (2018). Noncognitive proponents' conflation of "cognitive skills" and "cognition" and its implications. *Personality and Individual Differences, 134*, 25–32. https://doi.org/10.1016/j.paid.2018.05.025.

Kell, H. J. (2022). The criterion problem in cross-cultural performance research. *International Journal of Cross-Cultural Management, 22*(3), 389–411. https://doi.org/10.1177/14705958221100669.

Kell, H. J., & Lubinski, D. (2013). Spatial ability: A neglected talent in educational and occupational settings. *Roeper Review, 35*(4), 219–230. https://doi.org/10.1080/02783193.2013.829896.

Kell, H. J., Lubinski, D., Benbow, C. P., & Steiger, J. H. (2013). Creativity and technical innovation: Spatial ability's unique role. *Psychological Science, 24*(9), 1831–1836. https://doi.org/10.1177/0956797613478615.

Kell, H. J., McCabe, K. O., Lubinski, D., & Benbow, C. P. (2022). Wrecked by success? Not to worry. *Perspectives on Psychological Science, 17*(5), 1291–1321. https://doi.org/10.1177/17456916211055637.

Kell, H. J., & Wai, J. (2019). Right-tail range restriction: A lurking threat to detecting associations between traits and skill among experts. *Journal of Expertise, 2*(4), 224–242.

Klimoski, R. J. (1993). Predictor constructs and their measurement. In N. Schmitt, & C. Borman (Eds.), *Personnel selection in organizations* (pp. 99–134). Jossey-Bass.

Koenig, K. A., Frey, M. C., & Detterman, D. K. (2008). ACT and general cognitive ability. *Intelligence, 36*(2), 153–160. https://doi.org/10.1016/j.intell.2007.03.005.

Lakin, J. M., & Kell, H. J. (2020). Intelligence and reasoning. In R. J. Sternberg (Ed.), *The Cambridge handbook of intelligence* (pp. 528–552). Cambridge University Press. https://doi.org/10.1017/9781108770422.023.

Lakin, J. M., & Wai, J. (2020). Spatially gifted, academically inconvenienced: Spatially talented students experience less academic engagement and more behavioral issues than other talented students. *The British Journal of Educational Psychology, 90*(4), 1015–1038. https://doi.org/10.1111/bjep.12343.

Lam, S. S., Ng, T. W., & Feldman, D. C. (2012). The relationship between external job mobility and salary attainment across career stages. *Journal of Vocational Behavior, 80*(1), 129–136. https://doi.org/10.1016/j.jvb.2011.05.002.

Lubinski, D. (2000). Assessing individual differences in human behavior: "Sinking shafts at a few critical points". *Annual Review of Psychology, 51*, 405–444. https://doi.org/10.1146/annurev.psych.51.1.405.

Lubinski, D. (2009). Cognitive epidemiology: With emphasis on untangling cognitive ability and socioeconomic status. *Intelligence, 37*(6), 625–633. https://doi.org/10.1016/j.intell.2009.09.001.

Lubinski, D., & Benbow, C. P. (2000). States of excellence. *The American Psychologist, 55*(1), 137–150. https://doi.org/10.1037/0003-066X.55.1.137.

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*(4), 316–345. https://doi.org/10.1111/j.1745-6916.2006.00019.x.

Lubinski, D., & Benbow, C. P. (2021). Intellectual precocity: What have we learned since Terman? *The Gifted Child Quarterly, 65*(1), 3–28. https://doi.org/10.1177/0016986220925447.

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*(12), 2217–2232. https://doi.org/10.1177/0956797614551371.

Lubinski, D., Benbow, C. P., Webb, R. M., & Bleske-Rechek, A. (2006). Tracking exceptional human capital over two decades. *Psychological Science, 17*(3), 194–199. https://doi.org/10.1111/j.1467-9280.2006.01685.x.

Lykken, D. T. (1968). Statistical significance in psychological research. *Psychological Bulletin, 3*, 151–159. https://doi.org/10.1037/h0026141.

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*(7), 1004–1018. https://doi.org/10.1177/0956797616644735.

Makel, M. C., & Plucker, J. A. (2014). Facts are more important than novelty: Replication in the education sciences. *Educational Research, 43*(6), 304–316. https://doi.org/10.3102/0013189X14545513.

McCabe, K. O., Kell, H. J., Lubinski, D., & Benbow, C. P. (2018, July). Contributions of gifted individuals in enhancing their communities: A worldwide perspective [paper presentation]. In *Annual meeting of the International Society for Intelligence Research, Edinburgh, Scotland*.

McCabe, K. O., Lubinski, D., & Benbow, C. P. (2020). Who shines most among the brightest? A 25-year longitudinal study of elite STEM graduate students. *Journal of Personality and Social Psychology, 119*(2), 390–416. https://doi.org/10.1037/pspp0000239.

Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, sir Ronald, and the slow progress of soft psychology. *Journal of Counseling and Clinical Psychology, 46*, 806–834. https://doi.org/10.1037/0022-006X.46.4.806.

Messick, S. (1979). Potential uses of noncognitive measurement in education. *Journal of Education & Psychology, 71*(3), 281–292. https://doi.org/10.1037/0022-0663.71.3.281.

National Research Council. (2012). *Education for life and work: Developing transferable knowledge and skills in the 21st century*. National Academies Press.

Olszewski-Kubilius, P., Subotnik, R. F., & Worrell, F. C. (2015). Antecedent and concurrent psychosocial skills that support high levels of achievement within talent domains. *High Ability Studies, 26*(2), 195–210. https://doi.org/10.1080/13598139.2015.1095077.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*, aac4716. https://doi.org/10.1126/science.aac4716.

Park, G., Lubinski, D., & Benbow, C. P. (2007). Contrasting intellectual patterns predict creativity in the arts and sciences: Tracking intellectually precocious youth over 25 years. *Psychological Science, 18*(11), 948–952. https://doi.org/10.1111/j.1467-9280.2007.02007.x.

Peters, S. J. (2021). The challenges of achieving equity within public school gifted and talented programs. *The Gifted Child Quarterly, 66*(2), 82–94. https://doi.org/10.1177/00169862211002535.

Preckel, F., Schmidt, I., Stumpf, E., Motschenbacher, M., Vogl, K., Scherrer, V., & Schneider, W. (2017). High-ability grouping: Benefits for gifted students' achievement development without costs in academic self-concept. *Child Development, 90*(4), 1185–1201. https://doi.org/10.1111/cdev.12996.

Ramos, A., Lavrijsen, J., Linnenbrink-Garcia, L., Soenens, B., Vansteenkiste, M., Sypre, S., Boncquet, M., & Verschueren, K. (2023). Motivational pathways underlying gifted underachievement: Trajectory classes, longitudinal outcomes, and predicting factors. *The Gifted Child Quarterly, 67*(3), 179–197. https://doi.org/10.1177/00169862221132279.

Reynolds, M. R. (2013). Interpreting the *g* loadings of intelligence test composite scores in light of Spearman's law of diminishing returns. *School Psychology Quarterly, 28*(1), 63–76. https://doi.org/10.1037/spq0000013.

Rinn, A. N., Mun, R. U., & Hodges, J. (2022). *2020–2021 state of the states in gifted education*. National Association for Gifted Children and the Council of State Directors of Programs for the Gifted.

Sackett, P. R., Putka, D. J., & McCloy, R. A. (2012). The concept of validity and the process of validation. In N. Schmitt (Ed.), *Oxford handbook of assessment and selection* (pp. 91–118). Oxford University Press.

Sackett, P. R., & Yang, H. (2000). Correction for range restriction: An expanded typology. *The Journal of Applied Psychology, 85*(1), 112–118. https://doi.org/10.1037/0021-9010.85.1.112.

Shea, D. L., Lubinski, D., & Benbow, C. P. (2001). Importance of assessing spatial ability in intellectually talented young adolescents: A 20-year longitudinal study. *Journal of Education & Psychology, 93*(3), 604–614. https://doi.org/10.1037/0022-0663.93.3.604.

Snow, R. E. (1999). Commentary: Expanding the breadth and depth of admissions testing. In S. J. Messick (Ed.), *Assessment in higher education: Issues of access, quality, student development, and public policy* (pp. 133–140). Lawrence Erlbaum Associates.

Soto, C. J., Napolitano, C. M., Sewell, M. N., Yoon, H. J., & Roberts, B. W. (2022). An integrative framework for conceptualizing and assessing social, emotional, and behavioral skills: The BESSI. *Journal of Personality and Social Psychology, 123*(1), 192–222. https://doi.org/10.1037/pspp0000401.

Subotnik, R. F., Olszewski-Kubilius, P., & Worrell, F. C. (2011). Rethinking giftedness and gifted education: A proposed direction forward based on psychological science. *Psychological Science in the Public Interest, 12*(1), 3–54. https://doi.org/10.1177/1529100611418056.

Super, D. E., & Bachrach, P. B. (1957). *Scientific careers and vocational development theory*. Bureau of Publications, Teachers College, Columbia University.

Terman, L. M. (1925). *Mental and physical traits of a thousand gifted children. Genetic studies of genius* (Vol. 1). Stanford University Press.

Thorndike, R. L. (1949). *Personnel selection: Test and measurement techniques*. Wiley.

Tran, B. T. N., Wai, J., McKenzie, S. C., Mills, J., & Seaton, D. (2022). Expanding gifted identification to capture academically advanced, low-income, or other disadvantaged students: The case of Arkansas. *Journal for the Education of the Gifted, 45*(1), 64–83. https://doi.org/10.1177/01623532211063936.

U.S. Census Bureau. (2012a). Bachelor's degree attainment tops 30 percent for first time. Census Bureau reports 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. http://www.census.gov/hhes/socdemo/education/data/cps/2012/tables.html.

Vaillant, G. E. (2000). Adaptive mental mechanisms: Their role in a positive psychology. *The American Psychologist, 55*(1), 89–98. https://doi.org/10.1037/0003-066X.55.1.89.

Wai, J. (2012). The SAT needs to be harder. *Education Week*. https://www.edweek.org/teaching-learning/opinion-the-sat-needs-to-be-harder/2012/07.

Wai, J. (2014). Experts are born, then made: Combining prospective and retrospective longitudinal data shows that cognitive ability matters. *Intelligence, 45*, 74–80. https://doi.org/10.1016/j.intell.2013.08.009.

Wai, J., & Guilbault, K. M. (2022). Multidisciplinary perspectives and field strengthening questions for gifted education research. *High Ability Studies, 34(1), 39-59*. https://doi.org/10.1080/13598139.2022.2064269.

Wai, J., & Lakin, J. M. (2020). Finding the missing Einsteins: Expanding the breath of cognitive and noncognitive measures used in academic services. *Contemporary Educational Psychology, 63*, Article 101920. https://doi.org/10.1016/j.cedpsych.2020.101920.

Wai, J., Lubinski, D., & Benbow, C. P. (2005). Creativity and occupational accomplishments among intellectually precocious youths: An age 13 to age 33 longitudinal study. *Journal of Education & Psychology, 97*(3), 484–492. https://doi.org/10.1037/0022-0663.97.3.484.

Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Education & Psychology, 101*(4), 817–835. https://doi.org/10.1037/a0016127.

Wang, L., Zhang, Z., McArdle, J. J., & Salthouse, T. A. (2008). Investigating ceiling effects in longitudinal data analysis. *Multivariate Behavioral Research, 43*(3), 476–496. https://doi.org/10.1080/00273170802285941.

Watts, T. W., Bailey, D. H., & Li, C. (2019). Aiming further: Addressing the need for high quality longitudinal research in education. *Journal of Research on Educational Effectiveness, 12*(4), 648–658. https://doi.org/10.1080/19345747.2019.1644692.

Webb, R. M., Lubinski, D., & Benbow, C. P. (2007). Spatial ability: A neglected dimension in talent searches for intellectually precocious youth. *Journal of Education & Psychology, 99*(2), 397–420. https://doi.org/10.1037/0022-0663.99.2.397.

Widiger, T. A. (Ed.). (2017). *The Oxford handbook of the five factor model*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199352487.001.0001.

Wirthwein, L., & Rost, D. H. (2011). Giftedness and subjective well-being: A study with adults. *Learning and Individual Differences, 21*(2), 182–186. https://doi.org/10.1016/j.lindif.2011.01.001.