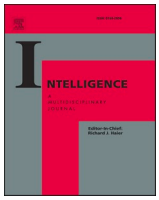
Intelligence 108 (2025) 101881

Contents lists available at ScienceDirect

Intelligence

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



Education, intelligence, placement, and selection: A discussion of paradoxes and fairness

David Lubinski

Vanderbilt University, United States of America

ARTICLE INFO

Keywords
Education
Cognitive abilities
Intelligence
Placement
Selection

ABSTRACT

Author's Note. This piece was initially prepared as a Foreword to *Testing and the Paradoxes of Fairness* by Howard Wainer and Daniel Robinson. However, in explicating the importance of that volume, and how further considerations only amplify its powerful argument, the amount of text required to do so became prohibitive. I am grateful to Wainer and Robinson for inspiring me to write this piece and especially to Richard Haier, Editor-in-Chief of *Intelligence,* who encouraged me to publish this piece in full in *Intelligence.*

1. Introduction

There is not a more pragmatically effective, yet maligned field in the behavioral sciences than the measurement of human abilities. It has been effective precisely because it has been built around measurement, and yet it is maligned precisely because it has been so effective in addressing real world issues. In *Testing and the Paradoxes of Fairness* Howard Wainer and Daniel Robinson masterfully describe how and why this has occurred.

The field of educational assessment, selection, and how learning and work opportunities are best distributed to maximize effective performance and minimize unfairness is critical if complex societies are to function effectively. It is hard to imagine a topic more in need of expert exposition at this time. Sadly, as important as this enterprise is to ensuring economic, personal, and physical well-being for all members of complex societies, it often engenders contentious debate. Discourse on how best to allocate precious educational-occupational resources frequently deteriorates into a toxic mix of name-calling and virtue signaling. Across boardrooms and university campuses, people walk on eggshells while navigating around this topic. All too often, the only dialog permitted is that devoid of robust, replicable empirical findings. Discussion is routinely insulated against invoking the powerful analytic tools that allow us to evaluate and measure desirable human outcomes, fatal human errors, and the neglected opportunity costs resulting from many kinds of poor decisions that lead to wasted efforts, misplaced resources, and dashed hopes. Unsecured credentials resulting from uninformed or misinformed educational decisions and the student debt crisis in the U.S. are among the many collateral casualties that result from ignoring research. Errors will always occur in human judgment and when developing and implementing public policies, but the authors of this book show us how to minimize dysfunctional outcomes and maximize beneficial ones. They do so in clean and compelling prose—scientifically informed and free of unnecessary jargon. Wainer and Robinson combine their vast expertise and historical knowledge with impeccable reasoning that hopefully will quell the contentious drama that surrounds appraisals of human potential by conceptualizing its conservation inclusively. And they do so with measured wisdom and grace.

To a degree, the friction in the contemporary landscape is understandable. Coveted positions in educational and occupational settings are intensely competitive. Resources are costly. Expertise requires years of study and training. Therefore, judicious procedures are required to select from the large pools of applicants by identifying accurately those most likely to persist with competence. The challenges posed by pandemics, cyber-insecurity, and climate change demand that those most competent and effective to address these issues are in the right places at the right time and are motivated to perform. At the same time, valid indicators of future learning and work performance, although predictively unbiased, do reveal demographic differences, thus ensuring that equity and fairness are ever present concerns. Many voices compete for the limited oxygen in this emotionally charged atmosphere. Such circumstances have given rise to privileging influences, which range from nepotism to group membership to entrenched ideological positions. Although there is a vast literature on the optimal allocation of precious human resources for learning and work—including civil service, government, and the military—it is complex and widely scattered across

E-mail address: david.lubinski@vanderbilt.edu.

https://doi.org/10.1016/j.intell.2024.101881
Received 26 November 2024; Accepted 26 November 2024

disciplines, journals, and technical reports. And everyone's time is precious. One reason why Wainer and Robinson's volume is remarkable is because it encapsulates in a short book what decades of sound research have taught us about managing and developing human potential in constructive and efficient ways. And in doing so, it reveals the many direct and indirect outcomes, either positive or negative, that arise from implementing novel policies and changing or maintaining existing ways of doing things.

Karl Popper (1959) argued that "*The main task of social science … is to trace the unintended repercussions of intentional human actions*" (p. 281, italics in original). Wainer and Robinson provide readers with a case study of why Popper's observation is so insightful. Without explicitly saying so, they show how Popper's maxim on unintended consequences can be combined with Lee Cronbach's (1983) recommendation for evaluating interventions or adjustments made to preexisting policies and programs (particularly those repeatedly validated as being effective). Their solution is as deceptively simple as it is compelling: collect data on multiple valued outcomes germane to all interested parties and make those data public. Wainer and Robinson provide multiple illustrations of the importance of doing so through positive examples of finding talent in economically-challenged and underserved communities as well as with tragic examples of the costs of neglecting empirical evidence. Topics covered range from pass/fail rates on law school bar exams as a function of matching/mismatching student readiness to curricula (Sander & Taylor, 2012), to scientific prerequisites and selection standards for medical training (Davis, 1986), to exposing cruel military practices that result from neglecting years of research on the learning and performance demands placed on past and modern combatants (Elliott, 1952, p. 83; Hamilton, 2015), and many others.

Outcomes should be evaluated not only in terms of the educational/occupational participants involved in selection but also their effects on society at large. We need both bridges and legal contracts that hold up when tested, as well as physicians who not only are able to pass licensing exams but also possess the personal attributes required for lifelong learning in a rapidly changing field. This is serious business. Indirect effects are ubiquitous. And the authors marshal poignant life and death examples that underscore why. Their humanitarian considerations generalize to the selection and training of master electricians and plumbers, air traffic controllers and commercial pilots, as well as law enforcement personnel. Although following Wainer and Robinson's wise counsel cannot resolve all the philosophical issues that surround how societies should best allocate opportunities or invest their resources, they clarify the benefits and costs associated with past and contemporary proposals for innovative practices. And they offer us a realistic way to foster consensus with empathy. Often those who initially disagree on the implementation of an educational or healthcare policy may come to a shared realization and agree on a particular way of doing things once the data are in and the human benefits, suffering, and opportunity costs are seen collectively, graphically, and quantitatively. Wainer and Robinson's suggested procedures also cast clarifying light on how to expose fraud and avoid policy incentives that tempt individuals into doing unwise things—as when 'suggestions' are made to *not* collect data. Collecting data to support one side's interests should not preclude other sides from doing the same for their viewpoints. Just as it isn't helpful to respond to pandemics by breaking all the thermometers, neither is doing away with valid tests for finding extraordinary potential among economically challenged communities. Doing away with valid indicators of valued human outcomes is never the right solution, however 'good' it might make us feel in the moment. Empirical findings do not *dictate* policy (Lees-Haley, 1996), but they can *inform* both experts and the public of the cascading direct and indirect costs and benefits likely to ensue by going down each of the many different paths. How best to assemble and assess consequential outcomes based on valid indicators of human individuality is what this volume is about, and it contains insights on outcomes seldom considered by the public or even professionals. Powerful causes and interventions have multiple effects. However, collecting data on multiple outcomes is necessary in order to see them clearly and holistically.

In the remainder of this discussion, I will suggest how and why the authors' perspective gains appreciable currency by considering some robust empirical findings and viewpoints on the assessment of general cognitive abilities that range across disparate sources published over the past century (Grove & Meehl, 1996; Kahneman et al., 2021; Meehl, 1990, 2006). While built on an exceedingly firm scientific foundation, I submit aspects of Wainer and Robinson's position are understated, especially those dealing with the outer envelope of human potential, and when viewed from an international perspective. When explicated, these considerations only reinforce the urgency of their views. When the authors say that "there is a bomb under the table," they are correct. There are huge individual differences in the growth rates of human proficiencies and managing them effectively is crucial for the well-being of students, workers, and our emerging global society. Neglecting these very real differences is done only at our peril. Therefore, let me then proceed to make that case.

2. A century of robust findings

Over a century ago, one of the co-founders of both the *Journal of Applied Psychology* (in 1917) and the *Journal of Educational Psychology* (in 1910) as well as the 20th president of the American Psychological Association (APA) published a prescient article in *Science*. Carl E. Seashore (1922) made known a fact that has withstood the test of time—namely, that the top 5 % of the freshman class at major U.S. universities is capable of consuming five times as much academic material as the bottom 5 % within the same amount of time. By contemporary standards, this was true for a remarkably homogeneous demographic during Seashore's era, the University of Iowa (which then generalized to peer institutions). A few years later, Hollingworth and Cobb (1928) published a stunning three-year longitudinal study of 8-year-olds. With ethnicity, gender, and socioeconomic status controlled, they compared the amount of learning displayed across two groups of students, all of whom were in the top 1 % in general intelligence test scores. One group was tightly clustered around an IQ score of 145, the other of 165. The amount of academic knowledge that these children acquired from age 8 to age 11 was markedly broader and deeper than that of their (typical) agemates. Importantly, these two groups also diverged conspicuously from one another across achievement assessments in Reading, Arithmetic Reasoning, and Science, as well as other subjects, commensurate with their IQ difference (with findings graphed to highlight their significance). Remarkably, the magnitude of the amount learned for the 165-IQ 11-year-olds was suppressed due to ceiling constraints on the outcome measures of academic knowledge; the full scope of their acquired knowledge could not be fully assessed due to the outcome assessments lacking a sufficient top end. Similarly, both criterion and predictor ceiling effects often operate to understate the findings cited by Wainer and Robinson throughout their volume.

The same year Hollingworth and Cobb published their classic study, Learned and Wood (1928) began their landmark studies of *The Student and His Knowledge*. They documented the vast range of individual differences in student knowledge both within and between grades and among high school students and those in college. They showed, for example, that 15 % of high school seniors ages 17 or less possess more general knowledge than the average college senior. These studies spawned a vast literature on the differential learning rates and educational needs found among students based on individual differences in learning readiness and the amount of time it took them to achieve outcome mastery (Paterson, 1957; Terman, 1954); Pressey's (1949) monograph is among the most compelling analyses of them. Learned and Wood's (1928, 1938) findings have been featured in classic texts on the psychology of individual differences for the past century (Anastasi, 1937; Tyler, 1965; Willerman, 1979). Their findings have been extended

using the specific abilities of mathematical, spatial, and verbal reasoning to make predictions about how specific strengths and relative weaknesses unfold into qualitatively distinct outcome profiles in educational → occupational → creativity and eminence over protracted intervals (Gohm et al., 1998; Humphreys et al., 1993; Lubinski & Benbow, 2021; Makel et al., 2016; Wai et al., 2009). For example, 12-year-olds assessed on college entrance exams such as the SAT mathematics and verbal reasoning subscales can consume a full high school course in 3 weeks' time if their scores are equivalent to the mean of college-bound high school seniors, while those with scores in the top 5 % can consume twice this amount (Benbow & Stanley, 1996; Stanley, 2000). Years later, these two scoring standards among 12-year-olds result in conferred doctorates at 10- and 20-times base rate expectations, respectively; these two scoring standards also ultimately give rise to more patents, publications, occupational stature, and income (Lubinski, 2016; Lubinski et al., 2023). Different selection criteria and contrasting interventions result in markedly different outcomes; however, multiple outcomes need to be assessed in order to see them. Even with ultimate educational degree held constant and prestige of the university attended controlled (Park, et al., 2008), creative outcomes assessed decades later become more likely at the higher ability regions found within the top 1 % of ability (which constitutes over one-third of the full ability range). Although many other factors matter, more ability is better because there is no ability-limiting threshold (Arneson et al., 2011). Since these findings are conspicuously neglected in the educational and psychological curricula—we had better ask, 'Why?'

An insidious and stubborn stereotype runs through modern educational practices, psychological theorizing, and policy development. Exposing this stereotype highlights one reason why this volume should be required reading for all educators, learning scientists, and public policy developers. The stereotype is that educational readiness is more uniformly tied to chronological age than it in fact is. There is, of course, some correlation between age and learning readiness, but as Binet discovered long ago, vast differences in "mental ability" exist *within* each chronological age. This holds true for both general and specific cognitive abilities, across all demographic groupings, and even within families (Lubinski, 2004). The demonstrated fact is that variation *between* groups is modest compared to variation *within* groups. And for optimal instruction, students need to be treated as individuals in terms of where they are at developmentally, not chronologically or demographically. That is why "time" is featured in John B. Carroll's (1963, 1989) compelling and empirically well-supported Model of School Learning. Due to vast differences in the amount of time required by students to master academic material—even among biological siblings reared together—optimal curriculum and instruction need to be individually tailored to maximize learning. This holds true for grades K-12 on through graduate *and* postdoctoral training. Ignoring this fact results in and explains the paradox of "Inequity in Equity" (Benbow & Stanley, 1996).

As the National Assessment of Educational Progress (NAEP) data reveal, 10 % of nine-year-olds are reading at or above the 25th percentile of 17-year-olds (Gottfredson, 2003). The same holds true for mathematical reasoning abilities. Such nine-year-olds have different learning needs. They also constitute a moving target because over the course of development, as Hollingworth and Seashore revealed—though this critical fact is ignored or even denied—education *expands*, rather than *reduces*, individual differences in achievement (Ceci & Papierno, 2005). In optimal learning settings while all students learn more, some learn more than others (Benbow & Stanley, 1996; Stanley, 2000). Some find this a conundrum that cries out to be "solved". Others find it distasteful and impolite—kind of like staring (and so they avoid it). And some view it simply as the nature of human individuality as in music or sports (Plomin, 2018). Regardless of one's perspective, failing to factor in this truism when developing curricula, evaluating selection, and conceptualizing inclusivity and fairness for all students is not only foolish, but wasteful of talent, and painful to many students forced to study in a 'one size fits all' classroom (Noreen et al., 2025).

Further, it is useful to take a modern cross-cultural perspective on education and advanced selection and training because this volume has international significance. Taking a global view highlights the strengths and weakness found within and between nations and cultures. While without doubt international allies and adversaries look with disbelief at many U.S. firearm policies, historically the U.S. has been viewed internationally as a model for advanced educational opportunities and graduate/postdoctoral training. For decades, the world has noticed and lamented the 'brain drain' as top U.S. universities attracted the best international minds available for their faculty and student body. Yet currently, some of the procedures for attracting and selecting exceptional talent have been called into question. Astonishingly, the very concept of merit itself has been called into question (Abbot et al., 2023) and jettisoning valid indicators of learning and work performance without empirical justification and/or the procedures to assess the direct and indirect effects of doing so has now become a common practice. In this context, the importance of collecting outcome data germane to all interested parties and making those findings public becomes more essential than ever. And here is one reason why.

Regardless of what educators and policy makers in the U.S. do, educational and talent development procedures across the globe are not tied to them. When Bill Gates built Microsoft Research Asia in Beijing, he systematically assessed a pool of 2000 PhD-level scientists on successively more demanding exams in programming and mathematics to winnow the pool down to 400 → 200 → 20. Ultimately, the top 20 were selected for two-year postdoctoral positions and, subsequently, 12 were permanently hired. He did the same the following year and thereby developed a world-class research institute in the same amount of time (*The World is Flat*, Friedman, 2005, p. 266). In *The Post-American World*, Fareed Zakaria (2011, pp. 205–206) detailed a similar procedure utilized in India. The entrance examination employed by India's Institute for Technology selects applicants at around the top 1 in 5000 (from a pool of 300,000). Both procedures are opaque in terms of the human individuality that they are indexing for selection, but modern psychological science has clarified the underlying dimensionality of the mathematical/spatial/verbal abilities being selected for (Kell et al., 2013; Makel et al., 2016). At all ages, the vast range of individual differences in reasoning with numbers, words, and shapes is as profound as it is consequential. Nowadays there are many people and psychological publications that enjoy talking about cognitive abilities and intelligence while routinely neglecting individual differences. However, they thereby sidestep or avoid predicting consequential human outcomes over protracted intervals. Wainer and Robinson's *Testing and the Paradoxes of Fairness* dives in where others do not dare dip their toes. They do so across the spectrum of talents found in the human condition as they build upon some of the most compelling insights and methods available in the social sciences.

Consider first, Daniel Kahneman et al.'s (2021) observation in "*Noise*", "[F]or all its crudeness and limitations, GMA [General Mental Ability], as measured by standardized tests containing questions on verbal, quantitative, and spatial problems, remains by far the best single predictor of important outcomes. … the predictive power of GMA is 'larger than most found in psychological research'" (pp. 229–230). Then next and from one of the originators of "Construct Validity" (Cronbach & Meehl, 1955), "Almost all human performance (work competence) dispositions, if carefully studied, are saturated to some extent by the general intelligence factor, *g*, which for psychodynamic and ideological reasons has been somewhat neglected in recent years but is due for a comeback" (Meehl, 1990, p. 124) … "its influence permeates all areas of competence in human life" (Meehl, 2006, p. 435). Just as Meehl identified ideological reasons for the neglect of general intelligence, Carroll (1989, p. 30) found the same bias to be responsible for neglecting the importance of time, which he sought to overcome in his Model of School Learning: "… objections to such practices stem largely from misplaced egalitarian attitudes. Equality of opportunity to attain potentials implies

that students with different amounts and kinds of aptitudes need to have educational programs that differ in pace and content, and perhaps many other ways. As someone has put it, we need not only equality of opportunity but diversity of opportunity." Leona E. Tyler, arguably the most distinguished counseling psychologist of the 20th century, made the same point 50 years ago.

> In our haste to abolish the unjust and the obsolete, we cannot afford to ignore the psychological realities that generated such systems in the first place. There are highly significant psychological differences among individuals, and the soundness of our social institutions depends upon how successfully we take them into account… A complex society cannot regard its members as identical interchangeable parts of a social machine. Its complex functioning depends upon the contributions of individuals specializing along different lines, equipped for carrying out different specialized tasks. For this reason, we must not be content with any system of universal education that provides identical treatment for all pupils. We must look for ways of diversifying education to make it fit the diverse individuals whose talents should be developed and utilized. (Tyler, 1974, pp. 6–7).

Vast ranges of learning readiness and human potential are found across all demographic groupings among individuals of similar ages. With a few (clinical) exceptions, there are no discrete categories of individual differences; there is typically a continuous spectrum of systematic variation. Demographic distributions of individual differences are highly overlapping, and their ranges are comparable. Moreover, atypical potentialities are typical because whatever is atypical for an individual is typical for a population. All parameters of potential and actualized competencies have atypical extremes. One reason why stereotyping is wrong—including ageism in learning readiness—is that it negates treating each person as an individual. Carroll (1989) is correct in saying that the failure to consider the time factor in learning is demotivating for students at the atypical extremes (Noreen et al., 2025). Doing so is insensitive to an empirical precept of human individuality. Further, learning opportunities that are insensitive to time likely attenuate the creative potential of those best equipped to meet the complex challenges facing our modern global society (Pressey, 1955; Park et al., 2013; Wai et al., 2010). Treating each person as an individual is where psychological assessment belongs in educational/occupational settings and indeed all settings. Assessing potential means taking time into consideration because the more potential one has the less time it takes to meet learning and performance standards. This is not only true for cognitive abilities but athletic and artistic ones as well. It's time to accept the importance of time.

Historically, these considerations have been stressed periodically in the ebb and flow of differential emphases on the importance of individual differences and their social significance (Hobbs, 1958). One of the most renowned psychoeducational statisticians of the past century, Quinn McNemar (1964) in his American Psychological Association (APA) Presidential Address threw down the psychological gauntlet by asking the question, "Lost: Our intelligence? Why?" In *Testing and the Paradoxes of Fairness*, Wainer and Robinson accept that challenge and highlight why the Elephant-in-the-Room of individual differences in learning readiness asked 60 years ago needs to be rediscovered.

Importantly, McNemar's penetrating analysis has recently been replicated on a large scale internationally (Pokropek et al., 2022). McNemar's rhetorical question was posed based upon the wisdom he likely gained from being a gifted farm boy educated in a one room schoolhouse who was afforded the flexibility for bright students to seek out books from all grades and to work from their level of knowledge already obtained (Hastorf et al., 1988). In his one room schoolhouse, McNemar worked his way through 8th grade mathematics by the end of 5th grade and the following year taught arithmetic to 8th graders (because the teacher was not competent to do it). When Wainer and Robinson detail Harvard's Admissions Officer Henry Chauncey (1930s) seeking out an economically impoverished student from the rural hills of Kentucky to populate the Ivy League, examples such as McNemar come to mind. Another example is Larry Hubert, given his unassuming early personal life history as well (Köhn et al., 2024); although when Hubert's teacher assigned him the task of 'helping out' his class, due to his advanced learning status, it definitely was not because the assigned instructor (Lee Cronbach) was incompetent to do so (Lubinski, 2016, pp. 934–935; Shavelson & Gleser, 2002).

There are many such diamond-in-the-rough examples to be found in the psychological literature, another stunning example being John E. Hunter (Schmidt, 2003), but as the authors stress, the plural of anecdote does not equal data. What does produce scientific data is when large samples are carefully studied across the full intellectual spectrum and over their lifespan. Doing so brings to light some spectacular positive outcomes (Makel et al., 2016) and shows how to avert a host of human tragedies (Hamilton, 2015). It also can expose discriminative practices, as in Henry Chauncey's role in correcting the selection policy of Harvard's President Abbott Lawrence Lowell (1920s), which had been intentionally designed to minimize the number of Jewish students admitted.

3. Test items, scales, measurement vehicles, psychological substance, and perspective

When constructing measures to assess human potential there are good reasons to focus on reasoning items in some contexts while items of knowledge prove better for others. Reasoning items that focus on discerning relationships between novel and familiar content are better for finding talent in rare places and are also better for mapping the outer envelope of intellectual functioning and detecting differential strengths in reasoning with words, numbers, and shapes. However, as the authors compellingly point out, comprehensive assessments of individual differences should examine knowledge and reasoning jointly as well as independently.

Perspective is needed because when comprehensive assessments of items exclusively focused on "reasoning" versus "knowledge" are conducted, they result in comparable measures or assessment vehicles for general purposes (Roznowski, 1987).

This also is what moved Cronbach (1976) to remark,

> *In public controversies about tests, disputants have failed to recognize that virtually every bit of evidence obtained with IQs would be approximately duplicated if the same study were carried out with a comprehensive measure of achievement* (p. 211, italics original).

Note that "comprehensive measure of achievement" is key. Hollingworth and Cobb's (1928) study used IQ whereas Learned and Wood (1928) used a comprehensive measure of general knowledge. It didn't matter for their purposes. Both assessment vehicles functioned equivalently for their general purposes. And indeed, when Truman Kelley (1927, p. 64) introduced measurement specialists to the Jangle Fallacy, the well-known practice of assigning different labels to experimentally independent scales that measure the same attribute (because psychologists can name more things than they can measure independently), he used broad measures of knowledge and reasoning to illustrate his point: "Equally contaminating to clear thinking is the use of two separate words or expressions covering in fact the same basic situation, but sounding different, as though they were in truth different. The doing of this … the writer would call the 'jangle' fallacy. 'Achievement' and 'intelligence' … We can mentally conceive of individuals differing in these two traits, and we can occasionally actually find such by using the best of our instruments of mental measurement, but to classify all members of a single school grade upon the basis of their difference in these two traits is sheer absurdity." APA's first Task Force on the distinction between aptitude and achievement tests arrived at the same conclusion (Cleary et al., 1975), which has withstood the test of time.

Intelligence has at least two important features. These two features are mirrored in items focused on knowledge versus those focused on reasoning. They were discussed by Thurstone (1924) when he distinguished between the products of intelligence (knowledge) and intelligence at work (discerning and effectively processing patterns and relationships across novel and familiar content). Within cultures, items assessing these "crystalized" versus "fluid" concepts are so inextricably intertwined that for many purposes item aggregates composed of broad samplings of each distills a communality that successfully and uniformly predicts the same pattern of external correlates. Yet, the item types are distinct. One way to distinguish intelligence as process from knowledge is to think of the forward versus reverse digit span. Essentially all adolescents know the number series 1 through 9, but there are huge individual differences in remembering a series of random numbers and then repeating them correctly in the same sequence (forward digit span). It is even harder to repeat them correctly in reverse order (reverse digit span). Both assessments modestly signal individual differences in general intelligence, however reverse digit span is the better (more valid, more highly correlated) predictor. And that is because it requires greater mental manipulation. Both tasks involve no new content but rather call for the manipulation of preexisting content. When a heterogeneous collection of such item types is professionally sampled and aggregated, their communality distills the construct of general intelligence ("*g*"). Spearman's (1927) underlying capacity is the ability to apprehend experience while educing relations and correlates across familiar and novel content (the essence of "*g*"); this attribute is the cornerstone upon which individual differences in acquiring knowledge are constructed (Corno et al., 2002; Snow, 1989, p. 22). Developmentally, acquiring knowledge and processing novel and familiar content more discerningly and efficiently unfold in parallel. When Carl Bereiter (1976/1977) said that intelligence is what you use when you don't know what to do, he had both components in mind.

Collectively, when it comes to acquiring knowledge, intellects are like snowballs rolling down a hill (Revelle et al., 2020). Some are stickier than others and pick up more volume as they traverse the vicissitudes of life. Parents see this among their own children. Knowledge readily and firmly sticks to some while others find it more elusive. That is why Woodrow (1921) said that intelligence is the capacity to develop capacities. Hollingworth and Cobb showed that exceptionally able children develop a greater volume of knowledge relative to able children. And the content of their knowledge is processed with differential degrees of nuance and swiftness. Impressive intellectual behaviors are associated with 'seeing' novel patterns in familiar things as well as storing knowledge. The individual differences that Hollingworth and Seashore described constitute the outer envelope of this capacity. And it is a capacity that runs continuously throughout the human condition (Gottfredson, 1997), cross-culturally (Pokropek et al., 2022; Warne & Burningham, 2019), and even comparatively (Flaim & Blaisdell, 2020; Lubinski, 2004). Importantly, this capacity has psychological penetrance outside of formal learning settings and occupational environments (Gottfredson, 2004; Underwood, 2014). Indeed, as Meehl (2006, p. 435) noted, "[I]ts influence permeates all areas of competence in human life."

Skilled psychometricians can systematically sample item types that vary in content and process to distill functionally equivalent indicators of the same general capacity to solve problems and reason as well as to acquire knowledge. Items of each type, whether reasoning or knowledge, individually carry tiny slivers of general intelligence ("*g*"). Systematically aggregating several of them distills their communality into a cohesive indicator of learning/performance capacity. Powerful constructs are carried by multiple vehicles. And different item types carry *psychological constructs* in different degrees and often in minuscule amounts. Yet, when aggregation across different item types is conducted, powerful constructs that possess the same functions are distilled by different vehicles. This property of test construction has been conceptualized by the term *systematic heterogeneity* (Hulin & Humphreys, 1980). Each individual item of reasoning and/or knowledge contains a tiny sliver of general factor ("*g*") variance even though it contains mostly construct-irrelevant noise. Yet, when professionals perform item samplings their communality is augmented while their specificities are attenuated, thus amplifying the signal-to-noise ratio of the composite. What is distilled though aggregation is primarily signal, with the specificities of each item (noise) reduced to tiny strands of construct irrelevancies having miniscule influence on the composite. The composite thereby consists of mostly what the items have in common ("*g*"). As Bert Green, (1978, p. 666) said "In Defense of Measurement," with properly conducted psychometrics, "given enough sow's ears you can make a silk purse." Wainer and Robinson's Fig. 9.1 in this volume dramatically depicts how quickly this can occur when item aggregates are amassed to distill appreciable commonalities (item intercorrelations averaging $r_{xx} = 0.15$ amass a communality of 0.90 when 50 systematically sampled items are summed).

Methodologically, the aggregation of distinct content that carries the same construct goes beyond systematic heterogeneity in test construction (Hulin & Humphreys, 1980). If a psychological construct is powerful, it *should* travel through multiple assessment vehicles—in this context different item and scale types. The fact that powerful constructs do so is the methodological workhorse underpinning such procedures as the Multitrait Multimethod Matrix (Campbell & Fiske, 1959), triangulating Converging Operations (Garner et al., 1956), minimizing Construct Irrelevancies in quasi-experimentation (Campbell & Stanley, 1963), Generalizability Theory (Cronbach et al., 1972; Humphreys, 1976), and orchestrating Constructive Replications (Lykken, 1968, 1991). These are powerful methodological procedures that all draw on essentially the same principle. Indicators of psychological attributes, constructs, or latent factors can never completely reduce unwanted noise or systematic sources of construct irrelevancies, but through aggregation of diverse items, measures, and scales they can appreciably reduce such extraneous influences so as to augment communality and thereby bring to light what they share. What Kahneman et al. (2021) refer to by the term "crudeness" in their positive appraisal of general intelligence, due to the "hotchpotch" of mathematical/spatial/verbal item types typically aggregated in unique ways to form General Mental Ability (GMA) composites, actually constitutes a methodological strength. When the construct irrelevancies in each item are minimized in the composite, what is distilled is a large communality of construct valid variance—the psychologically operative component of GMA. In their classic treatment of Unobtrusive Measures, Webb et al. (1966/1999, p. 29) captured the essence of this process by saying "there can be strength in converging weaknesses." Intellectual knowledge and cognitive processing operate in parallel and reflect individual differences that inextricably comingle. So, heterogeneous collections of both item types covary highly and generate the same pattern of external correlates for a variety of purposes.

At the outer envelope of intellectual challenges, extraordinary levels of both knowledge volume and the capacity to process both knowledge and novelty are needed to deal with unfamiliar circumstances. That also requires expertise composed of a highly specific knowledge base and procedural skills. When senior university administrators are confronted with a pandemic such as COVID-19 or the Apollo engineers are challenged by re-entry exigencies due to equipment malfunction, the content that they are dealing with is familiar—but it is the optimal rearrangement of that familiar content is not and so needs to be orchestrated in rapidly changing and unfamiliar circumstances. Developing a plan for safely navigating students and university personnel in their respective roles or for successfully guiding a re-entry vehicle safely to earth are merely two examples. That is why individuals need to be optimally placed in critical roles. Just as more able students are better at rearranging the content that all their peers have differentially mastered to generate effective solutions, more able professionals are more effective in responding to exigencies by rearranging familiar content in novel ways across unfamiliar settings. They have both more information to

draw on and they can process it more effectively in order to arrive at unique solutions in atypical circumstances. The ability to do so cuts across surgery, piloting, cybersecurity, rescue operations, and military intelligence. Spearman (1930, pp. 315–316) vividly illustrated this with the example of submarine officers engaged in battle as they deftly processed familiar content in novel ways to illustrate *g* acting operationally. That is why the measures, constructs, and methods presented in this volume are so important. They help us to conceptualize how to make optimal decisions.

This brief journey though the past century shows why John Stuart Mill's quote (which concludes this volume) on the ever-present soft whisper of something that is consistently true has merit. Over the past century, the authors' conclusions are supported by findings and methodological developments across many converging lines of evidence that address construing human potential and the circumstances required for it to flourish. Sometimes ideas need to be repeated by different people in different ways before they take hold—perhaps another example of the same information being carried by multiple agents, whereby aggregation is needed to quell the surrounding noise and so amplify the consensus developed over many years.

As for immediate exigencies, the authors are justified in warning us about that "bomb under the table." It is ticking loudly for those policy makers charged with developing procedures for allocating precious learning and work resources for general purposes and advanced training. There are no two more robust empirical generalizations in the applied psychological sciences than: (1) the extent to which general intelligence structures consequential outcomes in learning/work settings, and (2) the superiority of statistical over clinical prediction. Once experts have identified the best predictors for maximizing benefits and minimizing costs, it is time for statistical estimation not clinical conjecture (Grove & Meehl, 1996; Hubert & Wainer, 2013). Yet, both these forceful empirical generalizations have encountered intense and longstanding resistance. Several major universities have, without any supporting empirical evidence, jettisoned both practices (Abbot et al., 2023). One must therefore ask if evidence-based practice has any meaning for those institutions that have done so.

Finally, the content of this volume does more than simply clarify what is happening in contemporary applied psychoeducational assessment. It reveals why the "replication crisis" (Camerer et al., 2018; Open Science Collaboration, 2015) does not have to be a feature of contemporary social science, and it reinforces the reasons why longstanding uncontested epistemological and scientific principles for advancing knowledge are essential (Bleske-Rechek et al., 2024). In science, there are no "alternative facts," there are only facts. And when inconvenient findings are swept under the rug, they eventually come back to trip and to injure us. They impede the potential for scientific findings to alleviate human suffering and foster the identification and development of human potential. Although different individuals may assign different degrees of worthiness to the scientific measurement of human individuality, just like laws in the physical sciences, they operate whether or not we choose to measure them. Compelling scientific frameworks of human nature require the incorporation of all robust empirical findings. They require the assimilation of the Total Evidence (Carnap, 1950) without committing inferential fallacies by Neglecting Aspects (Castell, 1935) and embarking on Unprincipled Privileging (Oyama, 2000).

Testing and the Paradoxes of Fairness by Howard Wainer and Daniel Robinson is a model of how genuine theoretical insights in the social sciences are supported empirically and further refined. It is hard to imagine a theory of cognition or intelligence that holds prophecy for meaningful human outcomes that does not embrace the concepts, findings, and methods that they so ably marshal and communicate. In the words of Underwood (1975), individual differences provide the social sciences with "a crucible for theory construction." To theorize about cognitive abilities and intelligence in ways divorced from individual differences returns conceptual frameworks to the point where William James (1892, p. 335) left psychology long ago—with literary skill surpassing that of his brother Henry, "This is no science; it is only the hope of a science."

CRediT authorship contribution statement

David Lubinski: Writing – review & editing.

Declaration of competing interest

None.

Data availability

No data was used for the research described in the article.

References

Abbot, D., et al. (2023). In defense of merit in science. *Journal of Controversial Ideas, 3*, 1–26.

Anastasi, A. (1937). *Differential psychology*. Macmillan.

Arneson, J. J., et al. (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.

Benbow, C. P., & Stanley, J. C. (1996). Inequity in equity: How "equity" can lead to inequity for high-potential students. *Psychology, Public Policy, and Law, 2*, 249–293.

Bereiter, C. (1976/1977). IQ and elitism. *Interchange, 7*, 36–44.

Bleske-Rechek, A., et al. (2024). Behavioral science needs to return to the basics. *Skeptic, 29*, 28–37.

Camerer, C. F., et al. (2018). Evaluating the replicability of social science experiments in *nature* and *science* between 2010 and 2015. *Nature Human Behavior, 2*, 637–644.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81–105.

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook for research on teaching*. Rand McNally.

Carnap, R. (1950). *Logical foundations of probability*. University of Chicago Press.

Carroll, J. B. (1963). A model of school learning. *Teachers College Record, 64*, 1–9.

Carroll, J. B. (1989). The Carroll model: A 25-year retrospective and prospective view. *Educational Researcher, 18*, 26–31.

Castell, A. (1935). *A college logic: An introduction to the study of argument and proof*. Macmillan.

Ceci, S. J., & Papierno, P. B. (2005). The rhetoric and reality of gap closing: When the "have nots" gain but the "haves" gain even more. *American Psychologist, 60*, 149–160.

Cleary, A., et al. (1975). Educational uses of tests with disadvantaged students. *American Psychologist, 30*, 15–41.

Corno, L., et al. (2002). *Remaking the concept of aptitude: Extending the legacy of Richard E. Snow*: Erlbaum.

Cronbach, L. J. (1976). Measured mental abilities: Lingering questions and loose ends. In B. D. Davis, & P. Flaherty (Eds.), *Human diversity: Its causes and social significance* (pp. 207–222). Ballinger.

Cronbach, L. J. (1983). *Designing evaluations of educational and social programs*. Jossey-Bass.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281–302.

Cronbach, L. J., et al. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. Wiley.

Davis, B. (1986). Medical education and affirmative action. In B. D. Davis (Ed.), *Storm over biology: Essays on science, sentiment, and public policy* (pp. 163–201). Prometheus Books.

Elliott, R. M. (1952). Richard M. Elliott. In E. G. Boring, & G. Lindzy (Eds.), *Vol. IV. A history of psychology in autobiography* (pp. 75–95). Century Psychology Series.

Flaim, M., & Blaisdell, A. P. (2020). The comparative analysis of intelligence. *Psychological Bulletin, 146*, 1174–1199.

Friedman, T. L. (2005). *The world is flat: A brief history of the twenty-first century*. Farrar, Straus, & Giroux.

Garner, W. R., et al. (1956). Operationism and the concept of perception. *Psychological Review, 63*, 149–159.

Gohm, C. L., et al. (1998). Underachievement among spatially gifted students. *American Educational Research Journal, 35*, 515–531.

Gottfredson, L. G. (1997). Why *g* matters: The complexity of everyday life. *Intelligence, 24*, 79–132.

Gottfredson, L. S. (2003). The challenge and promise of cognitive career assessment. *Journal of Career Assessment, 11*, 115–135.

Gottfredson, L. S. (2004). Intelligence: Is it the epidemiologists' elusive "fundamental cause" of social class inequalities in health. *Journal of Personality and Social Psychology, 86*, 174–199.

Green, B. F. (1978). In defense of measurement. *American Psychologist, 33*, 664–670.

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical/statistical controversy. *Psychology, Public Policy, and Law, 2*, 293–323.

Hamilton, G. (2015). *McNamara's folly*. Infinity Publishing.
Hastorf, A. H., et al. (1988). Quinn McNemar (1900-1986). *American Psychologist, 43*, 196–197.
Hobbs, N. (1958). The compleat counselor. *Personnel and Guidance Journal, 36*, 594–602.
Hollingworth, L. S., & Cobb, M. V. (1928). Children clustering at 165 IQ and children clustering at 146 IQ compared for three years in achievement. *Teachers College Record, 29*, 3–34.
Hubert, L., & Wainer, H. (2013). *A statistical guide for the ethically perplexed*. Taylor & Francis.
Hulin, C. L., & Humphreys, L. G. (1980). Foundations of test theory. In E. G. Williams (Ed.), *Construct validity in psychological measurement* (pp. 5–12). Princeton, NJ: Educational Testing Service.
Humphreys, L. G. (1976). A factor model for research on intelligence and problem solving. In L. B. Resnick (Ed.), *The nature of intelligence* (pp. 329–339). Erlbaum.
Humphreys, L. G., et al. (1993). Utility of predicting group membership and the role of spatial visualization in becoming an engineer, physical scientist, or artist. *Journal of Applied Psychology, 78*, 250–261.
James, W. (1892). *Psychology: The briefer course*. Dover Publication.
Kahneman, D., et al. (2021). *Noise: A flaw in human judgment*. Little Brown Spark.
Kell, H. J., et al. (2013). Creativity and technical innovation: Spatial ability's unique role. *Psychological Science, 24*, 1831–1836.
Kelley, T. L. (1927). *Interpretation of educational measurements*. World Book.
Köhn, H. F., et al. (2024). Profiles in research: Lawrence J. Hubert. *Journal of Educational and Behavioral Statistics*. xx, xx.
Learned, W. S., & Wood, B. D. (1928). *The student and his knowledge*. Carnegie Foundation for the Advancement of Teaching.
Learned, W. S., & Wood, B. D. (1938). *The student and his knowledge*. Carnegie Foundation for the Advancement of Teaching.
Lees-Haley, P. R. (1996). Alice in validityland, or the dangerous consequences of consequential validity. *American Psychologist, 50*, 981–983.
Lubinski, D. (2004). Introduction to the special section on cognitive abilities: 100 years after Spearman's (1904) "'General intelligence,' objectively determined and measured.". *Journal of Personality and Social Psychology, 86*, 96–111.
Lubinski, D. (2016). From Terman to today: A century of findings on intellectual precocity. *Review of Educational Research, 86*, 900–944.
Lubinski, D., & Benbow, C. P. (2021). Intellectual precocity: What have we learned since Terman? *The Gifted Child Quarterly, 65*, 3–28.
Lubinski, D., et al. (2023). Composing meaningful lives: Exceptional women and men at age 50. *The Gifted Child Quarterly, 67*, 278–305.
Lykken, D. T. (1968). Statistical significance in psychological research. *Psychological Bulletin, 70*, 151–159.
Lykken, D. T. (1991). What's wrong with psychology anyway? In D. Ciccetti, & W. Grove (Eds.), *Thinking clearly about psychology* (pp. 3–39). University of Minnesota Press.
Makel, M., et al. (2016). When lightning strikes twice: Profoundly gifted, profoundly accomplished. *Psychological Science, 27*, 1004–1018.
McNemar, Q. (1964). Lost: Our intelligence? Why? *American Psychologist, 19*, 871–882.
Meehl, P. E. (1990). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it. *Psychological Inquiry, 1*, 108–141.
Meehl, P. E. (2006). The power of quantitative thinking. In N. G. Waller, L. J. Yonce, W. M. Grove, D. Faust, & M. F. Lenzenweger (Eds.), *A Paul Meehl reader: Essays on the practice of scientific psychology* (pp. 433–444). Erlbaum.
Noreen, G. D., et al. (2025). In their own voice: Educational perspectives from intellectually precocious youth as adults. *The Gifted Child Quarterly, 69*, xxx.
Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251). https://doi.org/10.1126/science.aac4716. Article aac4716.
Oyama, S. (2000). Causal democracy and causal contributions in developmental systems theory. *Philosophy of Science, 67*(Suppl), S332–S347.
Park, G., et al. (2008). Ability differences among people who have commensurate degrees matter for scientific creativity. *Psychological Science, 19*(10), 957–961.
Park, G., et al. (2013). When less is more: Effects of grade skipping on adult STEM productivity among mathematically precocious youth. *Journal of Educational Psychology, 105*, 176–198.
Paterson, D. G. (1957). The conservation of human talent. *American Psychologist, 12*, 134–144.
Plomin, R. (2018). *Blueprint*. Penguin Books.
Pokropek, A., et al. (2022). How much do students' scores in PISA reflect general intelligence and how much do they reflect specific abilities? *Journal of Educational Psychology, 114*, 1121–1135.
Popper, K. R. (1959). Critiques of classical theories of history. In P. Gardiner (Ed.), *Theories of history* (pp. 275–285). Free Press.
Pressey, S. L. (1949). *Educational acceleration: Appraisals and basic problems*. Ohio State University. Pressey, S. L. (1955). Concerning the nature and nurture of genius. *Scientific Monthly, 81*, 123–129.
Pressey, S. L. (1955). Concerning the nature and nurture of genius. *Scientific Monthly, 81*, 123–129.
Revelle, W., et al. (2020). Cognitive ability in everyday life: The utility of open-source measures. *Current Directions in Psychological Science, 29*, 358–363.
Roznowski, M. (1987). The use of tests manifesting sex differences as measures of intelligence: Implications for measurement bias. *Journal of Applied Psychology, 72*, 480–483.
Sander, R. H., & Taylor, S. (2012). *Mismatch: How affirmative action hurts students it's intended to help, and why universities won't admit it*. Basic Books.
Schmidt, F. L. (2003). John E. Hunter (1939-2002). *American Psychologist, 58*, 238.
Seashore, C. E. (1922). The gifted student and research. *Science, 56*(1458), 641–648.
Shavelson, R. J., & Gleser, G. (2002). Lee J. Cronbach (1916-2001). *American Psychologist, 57*, 360–361.
Snow, R. E. (1989). Aptitude-treatment interaction as a framework for research on individual differences in learning. In P. L. Ackerman, R. J. Sternberg, & R. G. Glasser (Eds.), *Learning and individual differences* (pp. 13–59). Freedman.
Spearman, C. (1927). *The abilities of man: Their nature and measurement*. Macmillan.
Spearman, C. (1930). Autobiography. In C. Murchison (Ed.), *Vol. I. A history of psychology in autobiography* (pp. 299–333). Clark University Press.
Stanley, J. C. (2000). Helping students learn only what they don't already know. *Psychology, Public Policy, and Law, 6*, 216–222.
Terman, L. M. (1954). The discovery and encouragement of exceptional talent. *American Psychologist, 9*, 221–230.
Thurstone, L. L. (1924). The nature of intelligence and ability (III). *British Journal of Psychology, 14*, 243–247.
Tyler, L. E. (1965). *The psychology of human differences*. Century.
Tyler, L. E. (1974). *Individual differences: Abilities and motivational directions*. Prentice Hall.
Underwood, B. J. (1975). Individual differences as a crucible in theory construction. *American Psychologist, 30*, 128–134.
Underwood, E. (2014). Starting young: Decades-old IQ test records from Scottish children have opened a unique window on how the brain ages. *Science, 346*, 568–571.
Wai, J., et al. (2009). Spatial ability for STEM domains: Aligning over fifty years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology, 101*, 817–835.
Wai, J., et al. (2010). Accomplishment in science, technology, engineering, and mathematics (STEM) and its relation to STEM educational dose: A 25-year longitudinal study. *Journal of Educational Psychology, 102*, 860–871.
Warne, R. T., & Burningham, C. (2019). Spearman's *g* found in 31 non-Western nations: Strong evidence that *g* is a universal phenomenon. *Psychological Bulletin, 145*, 237–272.
Webb, E. J., et al. (1966/1999). *Unobtrusive measures* (Vol. 2). Sage Publications.
Willerman, L. (1979). *The psychology of individual and group differences*. W. H. Freedman & Co.
Woodrow, H. (1921). Intelligence and its measurement: A symposium XI. *Journal of Educational Psychology, 12*, 207–210.
Zakaria, F. (2011). *The post-American world* (2nd ed.). W. W. Norton.