Predictive Measures for the High Value Decision Maker of the Future

by Jordan Peterson, Associate Professor of Psychology, University of Toronto

Every organization has its stated function. The degree to which the organization fulfills that function might be considered its productivity. Productivity might be considered in absolute terms (a certain amount of output per unit of input) or in relative terms (more or less productive than other similar organizations). Once the function has been specified, and the definition of productivity established, the question of efficiency immediately becomes paramount: how is it possible to get more for less, or to perform even better, comparatively? Perhaps there are two broad approaches to answering such a question.

The first might be regarded as sociological: improvements in system function can, in principle, lead to increased efficiency. The proponents of such an approach consciously or unconsciously accept certain axioms, with regards to productivity and its increase. "Human resources" are considered malleable raw material, of approximately equivalent initial potential. Their individual function, subsidiary to that of the group, can be substantively enhanced through the use of extrinsic motivation (salary, benefits, organizational power or status) and they can be both trained and managed. Such an approach is ultimately grounded in a worldview derived from John Stuart Mill and the intensely Protestant Puritans. Mill, who was intensely tutored by his father, believed that the human mind was a "blank slate," whose ultimate nature was determined by environmental circumstance (training, education, culture and extrinsic modification). The Puritans, for their part, believed that earthly success was a measure of heavenly favor (or that the worth of the soul could be judged by the success of the labor). The combination of these two belief sets allows for the following conclusion: managers, favored by God and circumstance, set the stage for the success or failure of their subordinates by appropriately or inappropriately manipulating the system that surrounds those subordinates. This role, which has a causal relationship to the success or failure of the organization, is sufficiently important to justify both high salaries and increased status, relative to subordinates, who are in turn easily replaced and retrained.

The second approach to solving the problem of enhanced productivity might be regarded as psychological: Improvements in *personnel selection* can, in principle, lead to increased efficiency. An entirely different set of axioms apply. "Human resources" are not merely raw material, and they are most definitely not of approximately equal initial potential. Instead, individuals differ widely in their intrinsic ability. The mind is not a blank slate and it is in large part biological differences that determine the relatively ability of various persons to function in complex and demanding environments. Thus, selection is of primary importance, training must necessarily play a secondary role, and the purpose of management is to set goals, to coordinate individually-motivated patterns of action, and to remove obstacles to progress.

Which of these approaches is most appropriate? It depends to some degree on the average environment from which individuals potentially in the organization might be drawn. Severe cultural and educational deprivation can reduce the most promising of individuals to something approximating a true blank slate, and little can be done once such a history has been instantiated. In the time of Mill, with vast gulfs of opportunity separating the rare fortunate from the illiterate masses, the philosophy of the blank slate was pragmatically tenable: the primary differences between individuals may have been a result of training. In the modern world, however, where the cognitive complexity of the average human environment exceeds the capacity of the even the most brilliant to encompass, the differences between people are increasingly a consequence of factors intrinsic to the person.

Given this, the question of efficiency and its enhancement can be differentiated into more specific questions: How much variability is there between people in terms of their individual productivity? What intrinsic psychological factors underlie or contribute to individual productivity? Can individual differences in productivity be predicted? What is the value, if any, of such prediction? Some of these questions have been answered, at least in part – but the residual value of answering them with ever increasing accuracy is so high that the attempt is exceedingly well justified.

In the 1990s a sequence of industrial/organizational studies broke new ground, conceptually and methodologically, with regards to the issue of productivity. First, Schmidt & Hunter (1998) and others (Gottfredson, 1996; Gottfredson, 1997a) demonstrated unequivocally that general intelligence (which is essentially the average of any random subset of tests of abstraction) was powerfully associated with a variety of real-world outcomes, accounting for approximately half of the variance in productivity in complex jobs. Concurrently, Hunter, Schmidt & Judiesch (1990) demonstrated (1) that individual workers in identical jobs varied in an economically important way in productivity and (2) that this variation increased as a function of the complexity of the job (standard deviation of productivity = 19, 32 and 48% of mean productivity for low, medium and high complexity jobs, respectively). This means that in low complexity jobs, the top 1% of workers produce 3X as much as the bottom 1%. For medium and high complexity jobs, respectively, the corresponding figures are 12X and 27X. The statistics indicated, furthermore, that the low performers in the latter category were not capable of doing the job at all (their productivity was calculated, "impossibly," at less than zero). Performance differentials are not restricted to the farthest extremes of the distribution, either: the top 1% of workers in high complexity jobs are also 127% more productive than the average workers. Hunter, Schmidt & Judiesch (1990) also considered the productivity differential obtaining between the top 1 and bottom 1% of the distribution in cognitive ability among a typical but hypothetical group of individuals employed in a high complexity job, and found that it approximated 5:1. What this all meant, theoretically, is that the individual differences in productivity between people are so high – particularly in complex jobs – that even small increments in selection accuracy might produce very large dollar value gains in output.

Schmidt & Hunter (1998) fleshed out this theoretical possibility by providing a formula for determining the dollar value of a given increment in predictive ability for a given organizational position. The formula contrasts current and potential predictors, and adjusts for wage (as an indicator of average dollar value productivity), job complexity and selection ratio (as the utility of a predictor increases with the number of people screened using that predictor. Their formula demonstrates that a predictive increment r = .20 (well below the average maximum obtained, for example, using personality tests) translated into a dollar value increase of more than $25,000 in productivity for a $75,000/year employee in a high complexity job (assuming a selection ratio of 1/10). Even an apparently tiny increment of r = .1 results in $13000 in heightened productivity. This demonstration had one immediate effect, and one long-term implication. The immediate effect was to revitalize the science of personality analysis. In 1968, Walter Mischel had sounded the death knoll for personality theorists by announcing that even the most statistically reliable trait measure could never account for more than a mere 9% of the variance in the output measure under consideration. At that time, however, the nature of effect sizes was not well understood, and r = .3 as a maximum (squared to obtain percentage of the variance accounted for) appeared trivial to the point of irrelevancy. However, as Schmidt and Hunter (1998) demonstrated (arguing along the same lines as Rosenthal (1991) and, later, Hemphill (2003)) correlation coefficients of less than .3 were not only not trivial, they were relatively large, within the social sciences (Hemphill, 2003) and outside of them (Rosenthal, 1991). According to Cohen (1988), who provided these estimates on theoretical grounds, a "small" effect size was .10, medium .30, and large .50. However, Hemphill's empirical study indicated that only 1/3 of published social science studies reported r's or r equivalents of .30 or above, and that only 5% of studies ever generated an effect approximating .5. The long term implication of such work is clear: organizations that maximize their ability to select will reap exceptional rewards for relatively minimal and admittedly inaccurate and error-prone predictive efforts.

Cognitive ability is not the only known predictor of productivity differences, either. A happy confluence of events – one of which is our increased knowledge of the nature of effect sizes – has brought personality testing back into the picture. In the last thirty years, personality psychometricians have converged theoretically and practically on a five dimensional model of human temperament and personality traits. These five dimensions include extraversion-introversion, a measure of positive emotion, gregariousness and social dominance, emotional stability-neuroticism, a measure of negative emotion and volatility, agreeableness-toughness, a measure of interpersonal sensitivity, warmth and empathy, conscientiousness-carelessness, a measure of persistence and integrity, and openness-traditionalism, a measure of aesthetic sensitivity and divergent thinking. Two of these traits have clearly been associated with enhanced productivity, in most if not all occupational/organizational situations: conscientiousness and emotional stability. The former may allow for predictions as high as $r = .3$, while the latter perhaps approximates .15 - .20. The other three traits can be somewhat usefully used to predict performance with regards to more narrowly defined parameters (openness, for example, is a reasonably powerful predictor of creativity).

As a consequence of these diverse bodies of research, we now know the answer to some of the differentiated questions posed earlier. First, there is immense productivity variation between individuals. Second, this variation can be predicted, using tests of general cognitive ability and personality, which assess stable, biologically predicated individual differences. Third, such prediction has substantive economic utility (indeed, it could be argued that the typical organization could do nothing remotely as costefficient and profit-enhancing as improving their psychometric selection procedures). Fourth – and, perhaps, less evidently – such prediction is so valuable and so accurate (1) that the search for additional incremental predictors is evidently justifiable (even with r's < .10) and (2) it could be argued that selection is demonstrably more effective than training, even when that training is sophisticated and intensive.[1] Finally, the second model of managerial action (select, guide and leave alone – except for the removal of obstacles) is more likely to be accurate than the first (lead, train, monitor and replace as necessary).

Individuals who wish to improve the efficiency and productivity of organizations are thus left with two broad options. The first is to perseverate under the illusion of managerial omnipotence and to continue to "improve" whatever is deemed dysfunctional at a systems level. The second and more justifiable is to develop ever-improved methods of selection and to consider improvement both in terms of predictive accuracy and pragmatic, within-organization implementability. It is this latter approach that this project is devoted towards.

To this end, our research group has been engaged in the development of a new generation of cognitive and personality tests. The classic psychometricians have been perfecting the assessment of general cognitive ability (IQ or general intelligence, g) for almost a century, and have provided efficient, reliable and valid measures. However, a parallel field of cognitive assessment, neuropsychology, emerged in the 1930's, and has been in development ever since – although such development has been limited to the field of medical/clinical neuropsychological evaluation. Neuropsychological tests are, in principle, associated with localized brain function in a way that IQ tests are not.

The brain evidently performs many functions, only a small portion of which are explicitly cognitive (as it regulates breathing, hormonal function, autonomic nervous system activity, temperature, motivation, and emotion). Of those cognitive functions, only a handful are devoted to the operations that seem most relevant to success in highly complex situations. Those are the functions traditionally attributed to the prefrontal cortex, a large section of the most recently evolved portions of the brain emerging, in phylogenetic development, from the motor cortex. The prefrontal cortex (PFC) – particularly the dorso-lateral section – allows percept, concept and action to be abstracted and manipulated prior to instantiation in behavior and assumption.

Standard general neuropsychological test batteries, such as the Halstead-Reitan and Luria-Nebraska, are outdated,[i] time-consuming, and difficult to administer. Furthermore, they have not been validated with the same degree of precision and certainty that typify many newer tests, and provide a low-resolution picture of PFC function, at best.[ii][ii] The newer tests, however, cannot generally be purchased (with a few minor exceptions) and have to be reconstructed painstakingly from oft-sketchy and fragmentary descriptions in clinical research reports. Furthermore, they have never been compiled into battery form – necessary, as a variety of tests must be utilized to produce a reasonably comprehensive picture of PFC function[iii] – and are certainly not available in an efficient and easy-to-administer format. We have in consequence developed a comprehensive, wide-ranging and easily usable computerized dorso-lateral prefrontal battery of such tests (see www.examcorp.com). (Transformation of tests from paper-and-pencil to computer format has a negligible effect on their validity,[iv] and may improve their reliability.[v]) Performance on this battery correlates r = .57 with averaged multiyear multi-observer job performance among managerial and administrative staff in the corporate world and predicts Harvard and University of Toronto academic grades over and above IQ and g). Pharmacists who do and do not make prescription errors can also be reliably differentiated by PFC performance, with the error-prone group performing on average at the 10th percentile of a standard management comparison group, and the normal group performing at the 45th percentile. Performance on this battery is also associated with officer-generated leadership ratings among enlisted men in the service of the US Navy and among US Naval Academy Midshipmen in their first year at the Naval Academy in Annapolis, Maryland. A complete description of this battery, which contains eight major components, is presented in Appendix 1 of this document.

The ExamCorp battery also contains a proprietary Big Five personality test, based on the adjective checklist of Goldberg. Our initial investigations using this test or variants thereof showed good predictive power for academic performance among university students at Harvard and at the University of Toronto, particularly with regards to conscientiousness, which was correlated with CGPA at r ~ .4. The Big Five test also worked reasonably well with factory-level workers (conscientiousness r ~ .2), but the factor structure and the predictive validity broke down in the case of individuals involved in high complexity jobs, who elevated their trait measures in a socially-desirable direction. In consequence, we have begun to experiment with a Big Five

Inventory that increases within-trait discrimination (breaking down each of the Big Five into their two most statistically evident subtraits) and that relies on a variety of forced-choice zero-sum methods to obtain theoretically unfakeable indices of trait personality. Individuals instructed to fake good on a standard Big Five scale and this new test have proved themselves capable of compromising the predictive validity of the former, but not the latter. We have also developed a measure of creativity, the Creative Achievement Questionnaire that shows substantive promise as a measure of creative production (rather than mere capacity).

In consequence, we propose broadly to

1. Develop the twin capacities to integrate IQ and prefrontal assessment using computerized technology that can be implemented at a distance. This means (a) generating and validating online tests assessing fluid and crystallized IQ in a manner that is reliable and valid and immune from cheating, and (b) integrating the results of such tests with the results of our current neuropsychological battery, streamlined and shortened, in such a manner that individual report generation is immediate, comprehensible and pragmatically acceptable to the end user.

2. Perfect the assessment and measurement of Big Five personality traits (particularly, but not exclusively, conscientiousness and emotional stability) in a manner that is focused on validity, rather than integrity of factor structure. This means continued testing and use of our current forced choice zero sum multitrait Big Five. We are also attempting to assess the relationship between Big Five traits and non-self report behavioral attributes and tendencies. Recently, for example, we have shown that extraversion and induced positive emotion increase the tendency for individuals to discount the future at an increased rate. We have also shown provisionally that high levels of neuroticism compromise cognitive function under stressful conditions (a demonstration that could account for the relationship between neuroticism and compromised job performance). We have also begun to develop a personality disorder screening instrument that further work should perfect to the point of industrial implementation.

3. Perfect our measures of creative production.

4. Integrate the cognitive ability assessment processes with the personality and creativity assessments, to provide a comprehensive evaluation instrument, appropriate for the screening and selection of applicants to academic environments and corporate organizations.

5. Finally, we wish to initiate work on the assessment of a factor that has not been sufficiently attended to with regards to the long-term prediction of success: breadth, accessibility and quality of social networks. Older individuals tend to perform more poorly on tests of fluid, but not crystallized intelligence – yet those who run corporations and major non-business organizations tend to be fifty or older (old enough for such cognitive deterioration to be well begun). This either means that the wrong people are running things, or that older people have additional advantages that a trait approach is missing. We think the most promising of these potential advantages is social positioning and access to the cumulative cognitive and economic resources that such positioning allows.

Additionally, we propose a sequence of immediate practical studies:

1. Screening of a complete incoming Rotman Business School cohort using computerized measures of dorsolateral function and zero-sum personality measures, as the initial stage of a longitudinal study, designed (1) to predict academic success and whatever other valid measures of competence might be appropriately generated at the Rotman and (2) to begin the process of assessment of potential for long-term corporate and general life success.

2. Immediately assess a representative and willing cross-section of Rotman alumni, classified by success, using the measures described previously.

Appendix 1: ExamCorp Dorsolateral Prefrontal Task Battery

All computerized tasks are completely mouse-activated, self-administered, and required no keyboard familiarity to complete. All tasks run on a standard Windows 98 equipped PC. The following tasks are included:

1. Random Object Span: Participants are asked to select a different object (left-hemisphere processed words/nameable pictures and right-hemisphere processed nonsense words/abstract pictures) each of 12 times from a randomly shape-shifting array of 12 objects. The score for this task was computed as the proportion of unique pictures clicked, divided by the time taken to complete the tasks. Similar tests assess mid-dorsolateral PFC [vi]; the "working-memory" capacity they assess has been integrally linked to reasoning power.[vii]

2. Acquired Spatial Association: Subjects are asked to associate each of 6 "cards" with each of 6 randomly distributed "lights". The participant's score is the total number of trials required to learn the associations (i.e., to choose the correct paired associate on 15 consecutive trials). Patients with right or left frontal lobe lesions (and right hippocampal deficits) show performance deficits on such tasks.[viii]

3. Random Letter Span: Participants are asked to randomize the letters in a particular span (say, randomize letters A-F). Performance was assessed as the maximum span successfully randomized. Frontally-damaged individuals perform poorly on such tasks.[ix]

4. Word Fluency: Participants are asked to generate as many words as possible beginning with a specific letter. The number of words generated over 5 minutes was used as the score. Proper nouns and alternative ending forms of words (…ing, …s) are not counted. Individuals with left frontal damage are impaired on fluency tasks.[x] PET scan indicates activation of the left dorsolateral PFC (area 46).[xi] Participants are asked to enter words with an onscreen mouse-keyboard of our design.

5. Recency Judgement: Participants are shown a string of objects (words, etc.) of variable length, then an object pair. For each trial, they are asked which of the two objects they saw most recently. The number of correct trials was taken as the score for this task. Frontally-damaged individuals do poorly on such tasks.[xii]

6. Sustained Attention (Continuous Performance): Participants are asked to mouse-click at the quasi-random appearance of a particular single object (target) during presentation of a string of "distractor" objects. The score is the proportion of targets clicked by the participant. The right PFC (as well as the right superior parietal cortex)

appears critically involved in sustained attention of this type,[xiii] regardless of sensory modality.[xiv]

7. Go/No-Go: Participants are asked to learn mouse-click response to one object, but inhibition of such response to another. This was scored as the number of trials required to learn the go/no-go rule (i.e., correct performance for 20 consecutive trials). Higher-order primates with prefrontal periarcuate lesions perform poorly on such tasks.[xv]

8. Response Inhibition Task: Participants are asked to mouse-click at the unpredictable appearance of a set of random numbers, but to inhibit that response when a particular specified number (the target) appears. The score is the ratio of correct responses to target (i.e., no mouse-click on target) to the total number of times the target appeared. Impulsive individuals tend to do poorly on tests such as these.[xvi]

[1] This is of course in part because individuals who are highly selected will benefit more from training – that is, they will learn faster and implement what they have learned more effectively.

---

[i] The Halstead-Reitan, for example, is more than forty years old.

[ii] Kolb B, Whishaw IQ. Fundamentals of Human Neuropsychology. New York: Freeman, 1980, pp. 453-458.

[iii] Boone KB, Ponton MO, Gorsuch RL. Gonzalez JJ, Miller, BL. Archives of Clinical Neuropsychology 1998, 13:585-595.

[iv] King WC, Miles, EW. Journal of Applied Psychology 1995, 80: 643-651.

[v] Russell EW. *Neuropsychological Review* 1995, 5:1-68.

[vi] Milner, Petrides & Smith, 1985; Petrides & Milner, 1982; Petrides, Alivisatos, Evans & Meyer, 1993; Wiegersma, van der Scheer & Human, 1990)

[vii] (Kyllonen & Christal, 1990)

[viii] (Petrides, 1985)

[ix] (Petrides et al. 1993; Wiegersma et al. 1990).

[x] (Damasio & Anderson, 1993)

[xi] (Frith, Friston, Liddle & Frackowiak, 1991)

[xii] (Milner et al. 1985)

[xiii] (Buchsbaum, Nuechterlein, Haier, Wu, Sicotte, Hazlett, Asarnow, Potkin & Guich, 1990)

[xiv] (Pardo, Fox & Raichle, 1991)

[xv] (Petrides, 1986)

[xvi] (Pliszka, Borcherding, Spratley, Leon & Irick, 1997)