

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

The Future of Testing

Buros-Nebraska Series on Measurement and
Testing

1986

2. Testing Old, Testing New: Schoolboy Psychology and the Allocation of Intellectual Resources

Gene V. Glass

University of Colorado at Boulder, gene.glass@colorado.edu

Follow this and additional works at: <https://digitalcommons.unl.edu/burosfuturetesting>

Glass, Gene V., "2. Testing Old, Testing New: Schoolboy Psychology and the Allocation of Intellectual Resources" (1986). *The Future of Testing*. 3.

<https://digitalcommons.unl.edu/burosfuturetesting/3>

This Article is brought to you for free and open access by the Buros-Nebraska Series on Measurement and Testing at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in The Future of Testing by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

2 Testing Old, Testing New: Schoolboy Psychology and the Allocation of Intellectual Resources

Gene V Glass

University of Colorado, Boulder

There will be no divining of the future here—no megatrends, no reference to Orwell (except this one). What little we truly know about the future does not bear mentioning. Nor shall I refer to micro-computers, data banks, and other gaudy paraphernalia that the future holds for us. The future is best seen in a rear-view mirror. We can at least hope to see more clearly the recent past we have traversed. Appearances of rapid change are usually superficial. If we see the past and present clearly, we will know as much of our future as it is ever permitted us to know.

Kenneth Boulding (1968) contends that the discovery of knowledge is absolutely unpredictable, since knowledge is the one thing that if we could predict when we would discover it we would have it already. The evolution of a technology, if it were not really radically new, might be predicted with some success; but then, if we can't predict the discovery of new knowledge and we can only predict changes in ordinary technology, then surely we have little idea of the important changes that lie ahead. New knowledge is revolutionary; technology is Establishment. The discovery of knowledge upsets things, changes the way lives are led. Technology serves old entrenched interests and established institutions. The glacial evolution of testing in this century reveals the source of its momentum—new technologies are moving it, not new knowledge. Testing is the conservative wing of the Social Science party.

A Point of View: Abstracted Empiricism

The most revealing perspective to assume for viewing the evolution and the current condition of testing is that which affords the clearest picture of how testing relates to the basic disciplines in the study of human behavior. In the last

100 years, testing has moved gradually from the center to the periphery of the behavioral and social sciences. Once an integral part of the best thinking on human development and behavior, testing has progressively grown more inbred and dissociated from the leading theoretical positions in psychology and the social sciences. In its position at the margin, testing has come to serve more faithfully the goals of its own professional subculture and of a particular political subculture (i.e., its own intellectual Establishment and the professional-managerial Establishment) than to serve the ends of science and the true aims of education. To fulfill its promise, testing must find its way back to the center of psychological thinking.

My message here does little more than echo a theme sounded by Anne Anastasi (1967) in her 1966 presidential address to Division 5 of the American Psychological Association.

. . . psychological testing is becoming dissociated from the mainstream of contemporary psychology. Those psychologists specializing in psychometrics have been devoting more and more of their efforts to refining the techniques of test construction, while losing sight of the behavior they set out to measure. Psychological testing today places too much emphasis on testing and too little on psychology. As a result, outdated interpretations of test performance may remain insulated from the impact of subsequent behavior research. It is my contention that the isolation of psychometrics from other relevant areas of psychology is one of the conditions that have led to the prevalent public hostility toward testing (p. 297). Although the very essence of psychological testing is the measurement of behavior, testing today is not adequately assimilating relevant developments from the science of behavior. . . .

It is noteworthy that the term "test theory" generally refers to the mechanics of test construction, such as the nature of the score scale and the procedures for assessing reliability and validity. The term does not customarily refer to psychological theory about the behavior under consideration. Psychometricians appear to shed much of their psychological knowledge as they concentrate upon the minutiae of elegant statistical techniques. Moreover, when other types of psychologists use standardized tests in their work, they too show a tendency to slip down several notches in psychological sophistication (p. 300).

Anastasi saw several reasons for this unfortunate dissociation of psychological measurement from psychological theory. Increasing specialization in all disciplines has lessened the chances that one individual will be conversant with both the technical rigmarole that has come to characterize modern psychometrics and the theories of psychology themselves. The expense involved in developing major tests militates against changing them; thus the tests of today reflect the psychology of yesterday. Psychometricians have capitulated to unrealistic public demands for short cuts and magic—psychological theory is often ravaged in the process.

Four years ago, also on the occasion of an American Psychological Associa-

tion Division 5 presidential address, Robert Glaser (1981) voiced the same concern expressed by Anastasi 15 years earlier: Testing is estranged from its roots in psychological theory. In the written version of his address entitled "The Future of Testing: A Research Agenda for Cognitive Psychology and Psychometrics," Glaser attempted to present ". . . areas of social concern in which education and testing might profit from coordination with potentially helpful areas of psychological research" (p. 935). Glaser applied his understanding of recent advances in cognitive psychology to a critique of existing school testing practices and pointed toward useful applications of recent developments in the psychology of learning and thinking. The work of Brown and Burton (1978) and of Siegler (1976) was suggested as a basis for ways of diagnosing failures in learning and intellectual performance. Herbert Simon's (Simon & Chase, 1973) imaginative research on the nature of expertise was advanced as a beginning in the assessment of differences in knowledge structures and cognitive processes between novices and experts. The work of Hunt, Sternberg, and others suggested to Glaser new views on the assessment of aptitudes with the ultimate goal of altering and building those abilities that early-day psychologists were prone to accept as being immutable.

My proposed point of criticism would seem *ad hoc* and unconvincing if it were said to apply somehow uniquely to the problems of measurement and testing. Fortunately, such is not so. C. Wright Mills (1959) argued forcefully that the schism between method and theory is everywhere evident in the social sciences. Aimless fact finding unguided by worthy conceptual analysis was christened "abstracted empiricism" by Mills. Once severed from worthwhile theoretical thinking, abstracted empiricism follows a bureaucratic course of development.

I doubt that I can advance any more helpful message than to commend once more to your attention the wisdom in the observations of my respected colleagues Anastasi and Glaser. It will give me the greatest satisfaction, in addition, if I can convince you to entertain an even broader scope of relevant psychological theory than they imagined as being a proper mooring place for psychological and educational measurement. But before making that attempt, permit me first to recount briefly how testing came to assume its current condition, which more people agree with each passing year is in need of repair.

Psychology and Psychometrics in 1980: The Golden Days

Testing as we have known it for the past 75 years was originally the tool of psychologists and social scientists living through the denouement of the great Western European empires. The social scientists of the first two decades of this century were, with varying degrees of consciousness, Social Darwinists. The cultural relativist anthropologists, such as Boaz and later Mead, are the excep-

tions that prove the rule. Regardless of one's contemporary political leanings, it is difficult today to read Galton or Terman without blanching at the coarse ugliness of it, e.g., Galton (1892) analyzing the genetic superiority of one (19th-century English) village over its neighbor. But enough has been written about this embarrassing era in the history of psychometrics (Block & Dworkin, 1976; Fallows, 1980; Gould, 1981), and I do not bring it up here again to heap insult on contumely.¹ I want instead to praise testing of that time; for in spite of its grossness then, it had something that it lost soon after and is missing today. At least the concern with measurement of human behavior in the early stages of its history was allied with the best thinking of the time on psychology and sociology, i.e., with evolution by natural selection, with the attempt to apply ideas of biological evolution to culture and society. This was more true, of course, of the European psychometricians than the Americans; as Sizer (1970, p. 15) observed: ". . . Americans engineered the idea of mental testing and adapted late nineteenth-century European theories to the realities of a more modern America. Terman, Thorndike, and the rest were pioneers, but more as engineers than as theoreticians." It would indeed be an act of insensitive second-guessing to think that Galton and Spearman and Goddard and Terman and the others were wrong *and* should have known better. They may well have been wrong, just as our best theories will seem puerile to future scientists, but they were the leading psychologists and social scientists of their day and testing was their most useful tool. It was, I submit, testing's golden era, and it has not known their like since. Perhaps, as with the triumphs of a precocious child, testing's early successes led to its current difficulties. The technology of testing was quickly wedded to the burgeoning field of statistical methods. The discipline began to grow specialized and esoteric. In the 1920s, testing entered a stage of hyper-rationalization from which it has never re-emerged. Multiple factor analysis plumbed the "vectors of the mind" (Thurstone, 1935) with machinery (centroids, tetrads, reference systems, etc.) beyond the ken of psychologist and social scientist. It is of more than passing significance that the increase in technical sophistication of the testing movement had little effect on tests themselves or the theories on which they were based. I cannot, for example, discern Thurstone or Holzinger's lineaments in any of the contemporary tests of intelligence. Indeed, the modern intelligence scale would seem a familiar artifact if placed in the hands of a suddenly reincarnated psychologist of the Edwardian period. I know of no science—save perhaps, anthropology or history—about which the same could be said.

The development of psychometrics from its early triumphs to the modern era parallels, peculiarly enough, the birth and growth of a bureaucratic agency—

¹For the most recent chapter in this historical controversy, see Snyderman and Herrnstein (1983), who argue that, in spite of the racist character of much of the psychometrics of the early part of this century, it is doubtful that the research directly influenced the passage of the Immigration Act of 1924. Their apologia reminds one of the man who, when accused of murdering three men and a dog, forthwith produced (before the court) in his defense the dog alive.

there are, indeed, lest anyone doubt it, bureaucracies and bureaucrats of ideas. It is a natural and human failing when one has access to and control over specialized information to exert that control against change (Selznick, 1953). Technological society's emphasis on expertise and specialization produce trained incompetence, the narrowing of the scope permitted for intellectual experimentation. Psychometricians, with control over arcane corners of mathematics, exemplify forces such as these (Andresky, 1972). As Robert Merton (1975) observed, the specialist begins to resist change because of vested interest in the current structure, of a society or culture, whether material or intellectual. "Adherence to the rules, originally conceived as a means, becomes transformed into an end-in-itself; there occurs the familiar process of displacement of goals, whereby an instrumental value becomes a terminal value . . ." (p. 28).

This displacement of goals took place in testing between about 1940 and 1960 as nearly as I can judge. Spurred by what we are told were the great victories for testing in World War II (Chase, 1948), all of which were atheoretical and pragmatic, the discipline of psychometrics began to take shape apart from psychology. It turned its back on psychology and reached instead for an independent, autonomous set of "principles of measurement" that transcended everything in particular. I remember my delight in discovering in 1960 that I could read with nearly complete comprehension the great treatise on psychometrics, Gulliksen's *Theory of Mental Tests* [1950], simply because I had overlearned college math and knew less than nothing at all about psychology.

The development of reliability and validity theory—two of the greatest achievements of the psychometric movement—can be viewed as an over-ambitious attempt to axiomatize a discipline. I view Cronbach's work over the past two decades (and in particular his collaboration with Meehl [Cronbach & Meehl, 1955]) as an attempt to correct the errant ways of methodologists who felt they could safely leave substance behind in the search for the abstract foundations of measurement. It took a while to drive home the point that the validity of testing is a quality of a complex use of information; it is not a property of a random variable. It has taken longer to make the point that reliability is no different, and just as a test has no validity, so it has no reliability either. Measurements have meanings, and they permit or obstruct thinking to various degrees. The process by which measurements are taken, as well as the ideas that gave rise to the measurements, are only judged in accord with how both—ideas and measurements—lead toward greater understanding. The relationship is reciprocal: constructs and observations, meanings and methods. The message is Cronbach and Meehl's.

By the early 1960s, it is fair to say that testing in psychology and education was severed from its roots in the study of human behavior. Indeed, testing flourishes today where the environment is starkly atheoretical (education) and withers in precisely those locales where thinking about human behavior is freshest and most exciting (psychology, psychiatry).

The Modern Testing Establishment

What then has become of testing and measurement? Estranged from the behavioral and social sciences, grown mathematically elaborate and worshipful of the “general principles of measurement,” what has become of testing since it reached maturity and autonomy in the academies of post-World War II America and Europe? In short, it sold out to the highest bidder; it went Establishment. In the world of work and in education, testing stepped forward to play the role of gatekeeper and management tool in the processing of human lives for the meritocracy: professional lightning rod attracting and grounding the anger of the excluded and discarded; factotum for society’s dirty work.

The testing industry and, more regrettably, the discipline of measurement in the Academy no longer serve science in its search for understanding nor education in its search to educe from individuals that which is best in them. Increasingly they serve a national and international system of processing people by number, managing the flow of bodies from institution to institution, documenting the expected progress of pupils through the vast educational “system.” For its efforts, the psychometric industry with its academic support system acquires huge caches of unexpended income.

The lack of articulation between measurement and substantive theory is particularly serious in education, that most pragmatic and atheoretical of all disciplines where testing is applied. Achievement test batteries are designed around what is thought to be the content of the school curriculum as determined by surveys of textbooks, teachers and other tests. Textbooks and curricula are designed, on the other hand, in part around the content of tests. One cannot discern which side leads and which follows; each side influences the other, yet nothing assures us that both are tied to an intelligent conceptualization of what an educated person ought to be.

Considering the prominence of testing in contemporary American schools, it is amazing to realize how useless testing seems to those closest to the core of schooling: teachers and pupils. It is scarcely any secret that teachers regard standardized ability and achievement tests as an irrelevance or worse. They complain that they learn nothing from the results that wasn’t obvious before the test was given; the scores give no clue as to what should be done to eradicate ignorance or take advantage of talent and skill. Testing is a transaction between the testing companies and school administrators, state education officials, government agencies, lawyers and other middlemen in the system of schooling.

Tests are not used by educators to decide how children should be educated because they are not designed for such purposes and are virtually worthless toward such ends (Hawkins, 1977). This fact greatly concerned Oscar Buros (1977) who decried the drifting away of achievement tests from what was taught in a course toward the goal of predicting individual differences in attainment at higher grades. After two decades of research on aptitude-treatment interactions,

it remains unclear whether one ought to teach so as to utilize the person's strongest aptitude, or teach so as to compensate for the weakest aptitude, or both or neither; and perhaps the meager harvest of so much research should be blamed on the primitive concepts of aptitude on which the tests are based. Glaser (1981) has complained for decades that tests are useless for deciding what it is that a child can and cannot do, hence the need to reference the scale that a test creates to the criterion of skills, knowledges, and understandings that comprise whatever it is that we think of as facility in reading, math, science and the like. The distortion of this eminently sensible call for "criterion referenced testing" (now repeated in hopes of productive hybridization of education and cognitive psychology [Glaser, 1981]) into item banks for behavioral objectives with cut-off scores for mastery is one of the more unfortunate inventions of modern psychometrics. My colleague David Hawkins at the University of Colorado reached back to Dewey in making a similar argument in his brilliant little book, *The Science and Ethics of Equality* (1977, p. 75):

. . . we need a framework of general ideas adequate to the developmental perspective in which all important abilities and talents should be viewed. This means that we should dig under the surface of those tests which have provided the empirical basis for so much of statistical psychometrics, and specifically the various intelligence tests. . . . I do not want to beat the IQ tests over the head. They are useful in their way, though as John Dewey said more than once, they are of little use to good teachers, who need both a refinement and an immediacy of discrimination in their daily work with children which global test averages do not provide.

Gullickson [1983] surveyed 30 education professors and 400 teachers to record their priorities for the content of educational measurement courses. Of 50 topics rated for desired emphasis in a college course, the greatest discrepancies between professors and teachers emerged on these topics: Teachers wished for "great emphasis" to be placed on ways of "interviewing pupils and parents," "observing pupils' work habits," evaluating "class discussions" and "interpersonal relationships"; the professors rated these topics as deserving only "slight emphasis." (In fact, these items were among the top ten rated items by the teachers and among the professors' bottom ten! The professors' highest rated topics were calculating the mean and variance and calculating correlation coefficients.)

The contemporary problem of "learning disabilities" is a case that can be advanced in behalf of the argument that testing in education follows the wrong lights and serves the wrong masters. Measurement of LD is virtually uniformly pursued across the United States today. IQ and achievement test scores are compared and "significant discrepancies" are tagged as evidence of LD (Shepard, Smith, & Vojir, 1983; Smith, 1982). The use of available published tests to do this work is encouraged by many factors: they are legally defensible, they

seem scientific and unimpeachable to parents, they are cheap and quick. But they are being used to measure what was called "under-achievement" three decades ago. Learning disability or lack of motivation? Can anyone tell the difference? Well, of course, but not in the naive and mechanical way that these notions are being translated into numbers by clinicians and psychometricians. It doesn't matter in the least whether one regards LD as "euphemizing" or not; the periodic purging of dysphemisms like "retardate" or "slow learner" would only seem unnecessary to one who had never been called such things. The whole idea of "learning disabilities" is that obstacles to learning are more variegated and worthy of detailed analysis than being ascribed to dullness or low intelligence. It is manifest that learning involves the hardware of the brain and of the communication channels; and when learning goes wrong, these as well as what a child has experienced in both his distant and recent history are implicated. From this point, the work of psychiatry, medicine, cognitive psychology and the like must begin, and what we know of the technology of measurement may have a contribution to make. But the absurdly premature translation of ill-formed concepts of LD into IQ versus achievement discrepancies was a managerial expedient of a particularly mindless sort, motivated by legal, political and professional interests. Unfortunately, it was all too typical of how these interests have used educational measurement.

Testing has found a new market in minimal competence testing in education and in licensure and certification in the workplace. I regard both of these with disappointment or disdain (Getz & Glass, 1979; Glass, 1978; Glass, 1979; Hogan, 1983; Olson, 1983). I need not go into details of these unseemly businesses. Perhaps it is enough to go on record again as believing that both applications of testing serve crass political ends: One, the extension of centralized political control of education; the other, the protection of economic self-interest in the workplace. Neither (and this goes for academic selection as well) is based on proof of utility that would justify its negative consequences in denial of opportunities or restriction of free trade.

Recently after reviewing and integrating the findings of over 500 psychotherapy outcome experiments, my colleagues and I (Smith, Glass, & Miller, 1980, p. 187) remarked thus on the state of the research art in this area:

Psychotherapy-outcome research lacks nothing by way of differentiated interventions. The literature on treatment is a veritable pharmacopoeia of prescriptions. The design of controlled experimentation has been refined to a science that is within the grasp of any researcher who owns a table of random digits and recognizes the difference between blind and sighted assessment. However, the measurement of outcomes seems to have been abandoned at a primitive stage in its development. Rating scales are thrown together with little concern expressed for their psychometric properties. Venerable paper-and-pencil tests . . . with roots planted vaguely in no particular theory of pathology or treatment are used to hunt for effects of short-term and highly specialized brands of psychotherapy. A superfluity of instru-

ments exists, and too little is known about them to prefer one to another. Little is known about their structure, and less is known about their sensitivity to treatment.

Thinking back to both of the Coleman et al. (1966; 1982) studies (of equality of educational opportunity and of public and private schools), the Follow-Through evaluation (House, Glass, McLean, & Walker, 1978) and more, I worry that modern testing and measurement—with their overweening attention to the pragmatic, the conventional, the traditional—have too often not merely failed to reveal the complexity and subtlety of human experience, but have actually denigrated the value of attempts to improve it. The telescoping of the immense variety in the Follow-Through models down into a standardized basic skills test and two dubious “affective” measures was a travesty for which the psychometric community might hold certain bureaucrats responsible (House et al., 1978), but the bureaucrats in question are quick to respond to such charges that they chose the measures from among the best that the psychometric arts had to offer.

BRINGING MEASUREMENT BACK TO PSYCHOLOGY: TOWARD A SOLUTION

Case Study Research

The growing significance of the naturalistic or “case-study” methodology in the social sciences poses important new problems for measurements; and if these problems are accorded the attention they deserve, the benefits may accrue not just to naturalistic methods but to measurement itself throughout the social sciences. The translation of experience into an observational record—everywhere the fundamental problem in measurement—always requires the imposition of some explanatory, theoretical structure. We tend to forget or ignore this fact in the established areas of testing and measurement, and then we accept the theoretical structure (bequeathed to our generation by faculty psychology or “self-concept” psychology) not as supposition but as reality. This bit of self-deception is more difficult to maintain in naturalistic research where experience is more complex and different theoretical systems still compete for favor. The naturalistic scene focuses our attention on several of the problems that need to be addressed by measurement theorists across the entire range of behavioral and social sciences: the necessary tie between theoretical structure and observation; the problems of not oversimplifying, of doing justice to the complexity of human systems (whether they be at the level of the individual, the group or a whole society); the difficulty of “slicing up the raw behavioral flux” and translating it into observations or numbers (Meehl, 1978); the unique problems attendant upon the use of human observers of human behavior (i.e., the problems of *Verstehen*

(Dilthey, 1959–68) or countertransference as behavioral science method as Devereux (1967) has written of it). These, I submit, are the key methodological problems that measurement must face if it is to further—rather than retard or play only a superfluous role in—the progress of the social sciences.

Theoretical Possibilities

While I take some comfort in knowing that my observations about what is wrong with testing today were seen earlier by my esteemed colleagues Anastasi and Glaser among other references, I shall derive more satisfaction if I can broaden ever so slightly the range of psychological theory to which testing and measurement would do well to attend. My colleagues have emphasized cognition and its role in learning. It is true that Anastasi mentioned the importance of attending to developments in personality theory, but she singled none out, advising her listener rather to keep abreast of relevant research in clinical and social psychology.

I wish to go further and counsel my fellow educationists and psychometricians that the most exciting and productive thinking outside the area of cognitive psychology is virtually untouched by psychometrics and that we could do less well than to turn our attention there when we seek the path back to the best scientific thinking on human behavior. The “affective domain,” as it is so inappropriately named, was not addressed by Anastasi or Glaser, and I take it as a favor that they have left it to me to extend their thesis into this challenging domain of human motives, desires, fears, wishes, antipathies, hopes, and frustrations.

Anastasi (1967) bowed in the direction of “personality” assessment in her 1967 address, but it seemed a hesitant gesture. Although she urged psychometricians to keep abreast of clinical psychology, her vagueness on the matter of what in particular in that vast area was worth attending to reflects another dissociation of theory and practice in American psychology, namely, the schism between psychology and psychiatry, or even the parallel split between academic and clinical psychology (splits that have their own political and intellectual roots).

Ever since psychiatry emerged from Bedlam and the scientific dark ages, the dominant theoretical perspective has been psychoanalytic, Freudian. Strangely, in spite of the antipathy with which psychodynamic theory is viewed by American academic psychologists, across the hall their colleagues in clinical psychology honor it as the pre-eminent theoretical system (Garfield & Kurtz, 1974, 1977). The extension of Freud’s work made by such investigators as Hartmann, Spitz, Jacobson, Mahler and now the younger generation of ego-psychologists has produced a most impressive and far-reaching psychological theory of human development and behavior (both “normal” and pathological). Permit me to declare myself. Coupled with neurophysiology and behavioral genetics, the psychoanalytic perspective represents our best hope for a comprehensive and useful account of the development and psychology of human beings. And yet, the link between psychoanalysis and psychometrics is virtually nonexistent.

Not only is any connection between psychoanalytic psychology and psychometrics impossible to discern, but the latter sometimes rather grandly imagines that it has disproved the former. A methodological critique or a factor analysis of Rorschach responses, or a horse race between statistical and clinical predictions will not deal with the challenge that neo-analytic ego psychology presents to empirical methodology. It is a curious fact unknown to nearly all who cite Meehl (1954) as the refutation of “clinical insight” (and then by implication “psychoanalytic theory”) that Paul Meehl’s theoretical leanings are self-proclaimed as psychoanalytic.² This is not the place to defend psychoanalysis against the charge that it is “unscientific”—a charge made about equally often as the charge that it is false, the two charges being contradictory, at least by Popper’s criterion of what constitutes a scientific proposition. The defense of the scientific status of psychoanalysis was presented some time ago in Hook (1959). I only wish to add here that it is scientific in precisely the most significant way and in the way in which too many psychological theories are inadequate, viz., the conclusions of psychoanalysis (e.g., the meaning of dreams, parapraxes, the operation of defense mechanisms, and the like) are “risky” propositions (in the Popperian sense), meaning that they are not independently derivable from common sense.

Psychoanalytic Psychology

It is my opinion that psychoanalytic psychology, particularly in its modern forms, represents the most significant challenge and opportunity that testing and measurement in the social and behavioral sciences could assume. It promises to change fundamentally the way in which psychometrics is pursued outside the narrow area of assessing aptitude and learning. If studied seriously, psychoanalytic psychology could lead to new techniques of observation and at least new concepts of the relationship between manifest behavior and the enduring psychology of the individual.

I see this possible relationship only vaguely myself. Perhaps an example or two will help bring these generalities into better focus.

The measurement of “self-concept” is one of the most active areas of psychometric concern, and yet the construct as embodied in modern tests is a hoary and naive thing scarcely developed any further than William James’s (1890) thinking nearly a century ago. In Wiley’s (1961) famous treatments of self-concept measurement, though she treats the theoretical foundations of the construct with respect, they are revealed to be little more than vague, commonsensical sketches of the conditioning of Philistine self-satisfaction by a pair of bland

²“I am confident that psychoanalytic concepts will be around after rubber band theory, transactional theory, attachment theory, labeling theory, dissonance theory, attribution theory, and so on, have subsided into a state of innocuous desuetude. . . . At the very least, psychoanalysis is an interesting theory, which is more than I can say about some of the ‘theories’ that are currently fashionable” (Meehl, 1978, p. 817).

parents molding little lumps of clay with praise and kisses. One current "theory," which is attracting psychometric attention (Marsh, Smith, & Barnes, 1983), is a seven-part model of the self-concept: The self-concept is hypothesized to comprise (a) the physical ability self, (b) the appearance self, (c) the peer relations self, (d) the parent relations self, (e) the reading self, (f) the mathematics self, and (g) the school subjects self. In a 10,000-word research report on this theory, one reads of factor analysis, multitrait/multimethod matrices, self-reports versus ratings by others, discriminant validity, divergent validity, halo bias, social desirability response sets, and on and on. About the only reference to psychological theory is, "An implicit assumption of most theorists is that self-concept is multifaceted" (p. 334). Further, it is said to be formed out of experience with the environment and interactions with "significant others." Clearly, this is a picture of psychometrics running amuck! Big methodological guns loaded with folk wisdom and truisms. This factor analytic mincing could go on forever; we could, of course, equally well discover "automobile self-concept," "favorite football team self-concept," or even "preferred Baskin-Robbins flavor self-concept."

The problem with this empirical hustle and bustle is that it is thoroughly innocent of any serious theory about the psychological sense of self: how it develops, what it is, how it can become sick, how to make it well. How adequately do the Rogerian theory and the other commonsense accounts of "self-concept" stand up to the "risky" tests of explanatory scope which alone will separate idle psychologizing from respectable theory, such risky tests as accounting for the ephemeral sense of identity of a thoroughly decompensated schizophrenic, the sense of gender identity so confused as to cause a man wilfully to mutilate his body surgically and chemically, the dangerous line between positively cathected self-representations and neurotic narcissism, the splintered selves of a multiple personality, or the more ordinary feelings of depression and emptiness in a child whose every act prompts nothing but praise and reward from the "significant others" of his environment? Unless our theories reach this far, they are in jeopardy of being not so much false as uselessly redundant with ordinary common sense.

It will surprise many to learn perhaps that one can scarcely find any reference in all of Freud's voluminous writings to a "self-concept," and that the term does not even appear in the lexicographic bible of psychoanalysis, Laplanche and Pontalis's *The Language of Psychoanalysis* (1973). Furthermore, in the writings of the neo-Freudians, those things that William James once thought of and the person-in-the-street now thinks of as the "self-concept" have been resolved into an extremely complex braid of developmental strands (including, among other things, identity formation *a la* Mahler) through stages of autism, merged self and object representations, differentiated representations, "practicing" and rapprochement subphases to gender identify and separation-individuation; or (*a la* Kohut) the formation of the ego ideal from disillusionment with the grandiose

self whose roots reach to the stage of infantile primary narcissism; or (*a la* S. Freud, A. Freud, Mahler, Jacobson and nearly every neo-Freudian) the construction of personal identity through identification with the loved or hated object. Neo-Freudian ego psychology has made exciting advances in those areas referred to colloquially as “self-concept” or “self-esteem”; the best didactic treatment of the past forty years of this research is the impressive two-volume work by Gertrude and Rubin Blanck, *Ego Psychology: Theory and Practice* (1974) and *Ego Psychology II: Psychoanalytic Developmental Psychology* (1979). This corpus of research can be commended to the attention of psychometricians; it has been attended to by at least one such, but more about that later.

In her summation on the state of self-concept measurement, Ruth Wiley (1961, pp. 317 ff.) criticized much of the psychometric work she reviewed on account of its theoretical naïveté. She was more generous than I might have been in the same situation (she being rather charitable toward some trivial conceptualizations), and her own grasp of the role of “self” in psychodynamic theory was weak. Nonetheless, she did identify the yawning gap between psychometric practice and psychoanalytic theory: “. . . certain psychologists have thought that self-concept research yields weak or equivocal results because the theory does not systematically include the unconscious self concept, or other unconscious cognitive and dynamic processes” (Wiley, 1961, p. 319). Although she thereafter goes on to place an unhealthy emphasis on the criterion of predictive validity for deciding whether new and unusual constructs (like the Freudian unconscious) should be allowed into the test battery, her sense of the seriousness of the omission of the unconscious from consideration of “self-concept” seems completely justified. Indeed, the situation is typical of academic psychology’s long relationship with Freudian psychology. Everywhere theorists and practitioners wish to cook the Freudian omelet without breaking the Freudian eggs. It cannot be done; the unconscious (whose existence is proved daily in our actions and nightly in our dreams) is the cornerstone of psychoanalysis and it cannot be locked in the closet like some shameful secret if psychology (and psychometrics) are to form any meaningful connection with psychoanalysis.

Consider Jane Loevinger’s (Loevinger & Wessler, 1970) work on the measurement of ego development. Loevinger, an early-day quantitative psychologist and psychometrician, has spent the last three decades engaged in an ambitious attempt to develop measures of some of those emotional-cognitive processes talked about by the neo-Freudians, and which she covers with the title “ego development.” Her efforts took off from a thorough understanding of traditional psychometric technique and its deficiencies for capturing what the ego psychologists were writing about: (a) there exists no one-to-one correspondence between a particular behavioral act and a level of ego development; (b) many strands of ego development occur simultaneously and one bit of behavior may reflect more than one strand; (c) no error-free way exists of distinguishing signs of developmental levels from signs of non-developmental correlates; (d) each individual

displays behavior at more than one level of ego maturity; (e) a behavioral sign may be discriminating in only one direction; and so on. Ironically, the one difficulty that Loevinger clearly identified and sought to overcome in her choice of psychometric format (sentence completion) and the voluminous and demanding scoring guides is precisely the point on which her critics allege that she departed from psychoanalytic theory. Loevinger wrote:

. . . no behavioral task can be guaranteed to display just what one wants to know about ego level. Neither a structured test nor an unstructured test carries a guarantee. If the test is structured, the investigator is projecting his own frame of reference rather than tapping the frame of reference of his subjects, which is the very thing that reveals their ego level. If the test is unstructured, one cannot control what the subject will choose to reveal (p. 9).

Loevinger must have felt she was steering a safe middle course through this dilemma by choosing incomplete sentences to be completed by the examinee (e.g., "When I get mad . . .," "When they avoid me . . .," "When they talked about sex, I . . .") and by struggling heroically with the free-form productions that result; but Gertrude Blanck (1976), whose understanding of Freud and the neo-Freudians is widely honored, blistered Loevinger for her misunderstanding of theory and for her choice of method: "The methodology . . . is simplistic, not alone by comparison (with psychoanalytic observational studies), but in its own right. Sentence completion as a research tool cannot be taken seriously because it relies on conscious responses and overlooks their unconscious determinants" (p. 803).

I can agree enthusiastically at a general level with what Mischel called for in his 1977 paper, "On the Future of Personality Measurement," viz., a broader assessment of persons functioning in their environment. I can endorse wholeheartedly and accept as my own his prediction that,

In the future, measurement hopefully will be directed increasingly at the analysis of naturally occurring behaviors observed in the interactions among people in real-life settings. . . . The future of personality measurement will be brighter if we can move beyond our favorite pencil-and-paper and laboratory measures to include direct observation as well as unobtrusive nonreactive measures to study lives where they are really lived and not merely where the researcher finds it convenient to look at them (p. 248).

Hear! Hear! And yet. . . . What kind of peeping does Mischel have in mind? And does he realize that if the dignity of the individuals involved is respected, they will ultimately be the source of information about their own lives and what it is like to live them; the assessment of personality might better resemble a psychiatric interview than a bugged room with one-way mirrors. Although Mischel seems to recognize this somewhat and says that personality measurement must

rely increasingly on self-reports, he finally (and disappointingly from my perspective) asserts his identity as an American academic psychologist and repudiates the unconscious, even suggesting obliquely that a couple of experiments and his own textbook have disproved its existence. Lord! Don't these psychologists ever sleep? And if they do, do they never dream?

A "Reflection on Schoolboy Psychology"

It is with regard to the role of unconscious processes that the relationship between psychoanalysis and psychometrics will be determined. If psychometrics continues to view the unconscious with arm's length suspicion as something unsavory or pathological or unscientific, then the opportunities for useful collaboration will be few.

It is, I believe, through a largely unconscious process of identification that we become truly educated. We grow to be like those we love. Our teachers give us an identity—not facts, not even a significant amount of whatever golden things lie at the highest level of Bloom's taxonomy. We are, each of us, living out lives that we took from someone else, someone we have loved and whose image we hold close by being what they were. Freud (1914) said as much in his "Some Reflections on Schoolboy Psychology."

. . . it is hard to decide whether what affected us more and was of greater importance to us was our concern with the sciences that we were taught or with the personalities of our teachers. . . . We courted them or turned our backs on them, we imagined sympathies and antipathies in them which probably had no existence, we studied their characters and on theirs we formed or misformed our own. They called up our fiercest opposition and forced us to complete submission; we peered into their little weaknesses, and took pride in their excellences . . . we can now understand our relation to our schoolmasters. These men, not all of whom were in fact fathers themselves, became our substitute fathers (p. 242).

There is in these instances of unconscious identification more about the true course of education,³ the way it shapes and molds and occasionally transforms us, than there is in all the behavioral objectives and mastery quizzes and standardized tests that ever were written. Children take more from the adults of their world than knowledge or training or even chromosomes. Through a process of identification, which springs largely from unconscious motives, they take a way of living that reaches to every corner of their lives. The dynamics of identification form some of the more interesting themes in that unstudied genre of liter-

³The true aims of education, as Michael Oakshott [1972, p. 40] characterized them, are ". . . initiation into the mysteries of a human condition; the gift of self-knowledge and of a satisfying intellectual and moral identity."

ature perhaps best called “teacher fiction” (e.g., the relationship of Godfrey St. Peter and Tom Outland in Willa Cather’s *The Professor’s House*, or the loathsome dynamics in Muriel Spark’s *The Prime of Miss Jean Brodie*). We must look in such places as these (the unconscious origins of identification, for one) with such peculiar instruments as interviews and freely flowing association and, yes, even dreams, if we wish to find the secrets of how children come to assume their individual adult forms.

There are two sides to this matter of the enduring impressions that education sometimes leaves. I spoke of one side, that seen by the marble. Now let the sculptor speak. Another man, Loren Eiseley, who after publishing *The Immense Journey* was acclaimed scientist and poet by the likes of Auden, once described⁴ the view from the lectern:

Now, for many years an educator, I often feel the need to seek out a quiet park bench to survey mentally that vast and nameless river of students which has poured under my hands. In pain I have meditated: “This man is dead—a suicide. Was it I, all unknowingly, who directed, in some black hour, his hand upon the gun?” “This man is a liar and a cheat. Where did my stroke go wrong?” Or there comes to memory the man who, after long endeavors, returned happily to the farm from which he had come. Did I serve him, if not in the world’s eye, well? Or the richly endowed young poet whom I sheltered from his father’s wrath—was I pampering or defending—and at the right or the wrong moment in his life? Contingency, contingency, and each day by word or deed the chisel falling true or blind upon the future of some boy or girl.

Ours is an ill-paid profession and we have our share of fools. We, too, like the generation before us, are the cracked, the battered, the malformed products of remoter chisels shaping the most obstinate substance in the universe: the substance of man. Someone has to do it, but perhaps it might be done more kindly, more precisely, to the extent that we are consciously aware of what we do—even if that thought sometimes congeals our hearts with terror. Or, if we were more conscious of our task, would our hands shake or grow immobilized upon the chisel?

I do not know. I know only that in these late faint-hearted years I sometimes pause with my hand upon the knob before I go forth into the classroom. I am afflicted in this fashion because I have come to follow Dewey in his remarks that “nature is seen to be marked by histories.” As an evolutionist I am familiar with that vast sprawling emergent, the universe, and its even more fantastic shadow, life. Stranger still, however, is the record of the artist who creates the symbols by which we live. As Dewey has again anticipated, “No mechanically exact science of an individual is possible. An individual is a history unique in character.” “But,” he remarks, “constituents of an individual are known when they are regarded not as qualitative, but as statistical constants derived from a series of operations” (1962, p. 25).

⁴In a book dedicated to Letta May Clark, Eiseley’s English teacher at University of Nebraska High School: *The Mind As Nature*, No. 5 in the John Dewey Society Lectureship, New York, Harper & Row, 1962.

Eiseley went on to examine creativity: “. . . that enigma to which the modern student of educational psychology is devoting more and more attention” (1962, p. 28). In 1962 when Eiseley wrote these words, educational psychologists did indeed aspire to measure and explore such “statistical constants” as creativity and motivation and trust and perseverance and “social competence.” I am old enough to remember when psychometricians spoke unashamedly of their aspirations to capture more of human experience than the IQ. Now they seldom confess such ambitions, content instead, it seems, to refine endlessly the mathematical foundations of measuring nothing in particular. And this concerns me more than anything else about the present condition of testing and its future: that its disciples no longer share any sense of wonder or fascination about the development of thought, the nurture of talent, the mysteries of human personality. Without the aspiration to understand human growth and behavior, testing will drift further from those sciences that keep such aspirations alive.

REFERENCES

- Anastasi, A. (1967). Psychology, psychologists and psychological testing. *American Psychologist*, 22, 297–306.
- Andresky, S. (1972). *Social sciences as sorcery*. London: Deutsch.
- Blanck, G. (1976). An ambitious undertaking. *Contemporary Psychology*, 21, 801–803.
- Blanck, G., & Blanck, R. (1974). *Ego psychology: Theory and practice*. New York: Columbia University Press.
- Blanck, G., & Blanck, R. (1979). *Ego psychology II: Psychoanalytic developmental psychology*. New York: Columbia University Press.
- Block, N. J., & Dworkin, G. (Eds.) (1976). *The IQ controversy*. New York: Pantheon Books.
- Boulding, K. E. (1968). Expecting the unexpected: The uncertain future of knowledge and technology (pp. 158–175. In K. E. Boulding, *Beyond economics: Essays on society, religion and ethics*. Ann Arbor: University of Michigan Press.
- Brown, J. S., & Burton, R. R. (1978). Diagnostic models for procedural bugs in basic mathematical skills. *Cognitive Science*, 2, 155–192.
- Chase, S. (1948). *The proper study of mankind*. New York: Harper and Row.
- Coleman, J. S., Campbell, E. A., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of educational opportunity*. Washington, D.C.: U.S. Government Printing Office.
- Coleman, J. S., Hoffer, T., & Kilgore, S. (1982). *High school achievement: Public, Catholic and private schools compared*. New York: Basic Books.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281–302.
- Devereux, G. (1967). *From anxiety to method in the behavioral sciences*. The Hague: Mouton & Co.
- Dilthey, W. (1959–68). *Gesammelte Schriften*. 14 volumes. Stuttgart: Teubner.
- Eiseley, L. C. (1962). *The mind as nature*. New York: Harper and Row.
- Fallows, J. (1980, Feb.). The tests and the ‘brightest.’ *Atlantic*, pp. 37–48.
- Freud, S. (1914). Some reflections on schoolboy psychology (pp. 241–250, Vol. 13). In *Totem and taboo and other works (1913–14)*. *Standard Edition of the Complete Psychological Works of Sigmund Freud*. James Strachey (Ed.), 24 vols. London: Hogarth Press, 1953–66.

- Galton, F. (1892). *Hereditary genius: An inquiry into its laws and consequences*. Bath, England: Pitman Press.
- Garfield, S. L., & Kurtz, R. (1974). A survey of clinical psychologists: Characteristics, activities and orientations. *The Clinical Psychologist*, 28, 7–10.
- Garfield, S. L., & Kurtz, R. (1977). A study of eclectic views. *Journal of Consulting and Clinical Psychology*, 45, 78–83.
- Getz, J., & Glass, G. V. (1979). Lawyers and courts as architects of educational policy. *High School Journal*, 62, 181–186.
- Glaser, R. (1981). The future of testing: A research agenda for cognitive psychology and psychometrics. *American Psychologist*, 36, 923–936.
- Glass, G. V. (1978). Standards and criteria. *Journal of Educational Measurement*, 15, 237–261.
- Glass, G. V. (1979, May 8). The war on incompetence. *Paper presented at the Distinguished Scholars Series*, Oregon College of Education; Monmouth, Oregon.
- Gould, S. J. (1981). *The mismeasure of man*. New York: Norton.
- Gullickson, A. R. (1983). Personal communication.
- Gulliksen, H. O. (1950). *Theory of mental tests*. New York: Wiley.
- Hawkins, D. (1977). *The science and ethics of equality*. New York: Basic Books.
- Hogan, D. B. (1983). The effectiveness of licensing. *Law and Human Behavior*, 7, 117–138.
- Hook, S. (Ed.) (1959). *Philosophy, scientific method and psychoanalysis*. New York: New York University Press.
- House, E. R., Glass, G. V., McLean, L. D., & Walker, D. (1978). No simple answer: A critique of the Follow-Through evaluation. *Harvard Educational Review*, 48, 128–160.
- James, W. (1890). *Principles of psychology*. New York: Holt. 2 volumes.
- Kamin, L. J. (1974). *The science and politics of IQ*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Laplanche, J., & Pontalis, J. B. (1973). *The language of psychoanalysis*. New York: Norton.
- Loevinger, J., & Wessler, R. (1970). *Measuring ego development, I*. San Francisco, CA: Jossey-Bass.
- Marsh, H. W., Smith, I. D., & Barnes, J. (1983). Multitrait-multimethod analyses of the self-description questionnaire: Student-teacher agreement on multidimensional ratings of student self-concept. *American Educational Research Journal*, 20, 333–357.
- Meehl, P. (1954). *Clinical vs. statistical prediction*. Minneapolis: University of Minnesota Press.
- Meehl, P. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, 48, 808–834.
- Merton, R. K. (1957). *Social theory and social structure*. New York: Macmillan.
- Mills, C. W. (1959). *The sociological imagination*. New York: Oxford University Press.
- Mischel, W. (1977). On the future of personality measurement. *American Psychologist*, 32, 246–254.
- Oakeshott, M. (1972). Education: The engagement and its frustration (p. 19–49). In R. F. Dearden, P. H. Hirst, & R. S. Peters (Eds.), *Education and the development of reason*. London: Routledge & Kegan Paul.
- Olson, P. A. (1983). Credentialism as monopoly, class war, and socialization scheme. *Law and Human Behavior*, 7, 291–299.
- Selznick, P. (1953). *The theory of social and economic organization*. New York: Oxford University Press.
- Shepard, L. A., Smith, M. L., & Vojir, C. P. (1983). Characteristics of pupils identified as learning disabled. *American Educational Research Journal*, 20, 309–331.
- Siegler, R. S. (1976). Three aspects of cognitive development. *Cognitive Science*, 8, 481–520.
- Simon, H. A., & Chase, W. G. (1973). Skill in chess. *American Scientist*, 61, 394–403.
- Sizer, T. R. (1970). Testing: America's comfortable panacea (p. 14–21). In G. V. Glass (Ed.), *Proceedings of the 1970 invitational conference on testing problems*. Princeton, NJ: Educational Testing Service.

- Smith, M. L. (1982). *How educators decide who is learning disabled*. Springfield, IL: C. C. Thomas.
- Smith, M. L., Glass, G. V., & Miller, T. I. (1980). *The benefits of psycho-therapy*. Baltimore, MD: John Hopkins University Press.
- Snyderman, M., & Herrnstein, R. J. (1983). Intelligence tests and the Immigration Act of 1924. *American Psychologist*, 38, 986–995.
- Thurstone, L. L. (1935). *The vectors of mind*. IL: University of Chicago Press.
- Wiley, R. C. (1961). *The self concept: A critical survey of pertinent research literature*. Lincoln: University of Nebraska Press.

