Daniel J. Keyser, Ph.D.
Richard C. Sweetland, Ph.D.

General Editors

TEST CRITIQUES
Volume VIII
John A. Johnson, Ph.D.
Associate Professor of Psychology, Pennsylvania State University,
DuBois Campus, DuBois, Pennsylvania.

PDI EMPLOYMENT INVENTORY
George E. Paajanen. Minneapolis, Minnesota: Personnel
Decisions, Inc.

Introduction

The PDI Employment Inventory (EI) is a screening device designed to predict whether prospective employees will be productive or counterproductive. Paajanen describes productive workers as those who show reliability, good work habits, and a motivation to conform to company policies. In contrast, counterproductive employees demonstrate irresponsibility, risk taking, instability, and dishonesty. The EI therefore can be conceptualized as an inventory assessing personality dispositions related to desirable work behavior (cf. Sackett, Burris, & Callahan, 1988).

Dr. George E. Paajanen is a graduate of the University of Minnesota and is currently Director of Products Development at Personnel Decisions, Inc. (PDI). Paajanen's (1988) dissertation compared the relative impact of personal dispositions and organizational variables on productive and counterproductive work behavior. The dissertation describes the development and validation of the EI and is the primary reference source for this review.

The development of the EI began with the identification of 25 counterproductive personality dispositions cited in the clinical and industrial psychology literature. An examination of 78 existing scales that seemed to assess at least one of these personality dispositions guided the composition of approximately 400 new test items. Additional items were written to measure characteristics related to short job tenure and to develop two validity scales—Frankness and Infrequency. Five industrial psychologists screened the items until 223 remained; these became the experimental EI.

Two samples completed the experimental EI. The first consisted of 4,652 persons applying for work at one of 81 Target discount department stores between December 1984 and June 1985. The EIs were returned without being scored at the interview site and were not used in hiring decisions. Within the 7-month testing period, 2,988 applicants were hired, 2,661 of which had complete EIs that could be compared with their personnel records. According to records, the sample was 38% male, 58% female, and 4% not designated. The racial composition of the sample was 73% white, 11% black, 7% Hispanic, 1% Oriental, 1% Native American, 1% other, and 9% not designated.

The second sample consisted of 109 University of Minnesota undergraduate students, and three additional measures were gathered that served as criterion data. This sample completed, in addition to the EI, Gough's (1972) Personnel Reaction Blank (a measure of conscientiousness and dependability), and a 20-item Legal Activities questionnaire (a form for confessing involvement in illegal activities). The students' honesty also was assessed by paying subjects who agreed to retake the EI $3.00 in advance and noting who defaulted on the agreement. Of 98 students indicating that they would retake the EI, 71 complied.

Item statistics were used to retain 97 items constituting four separate scales for the final version of the EI. The 52 items retained for the Performance scale (the principle scale assessing productivity potential) had to meet three criteria. First, items were required to show different response rates in the expected direction between 66 employees fired for gross misconduct and 654 persons employed at least 3 months with no disciplinary incidents. Second, items had to correlate significantly with either a performance rating in the employee sample or with one of the three criteria (dependability test, illegal activities, or defaulting for $3.00) in the student sample. Third, items had to demonstrate significant test-retest stability. Prospective employees will be productive or counterproductive workers who completed the test a second time 4 weeks after the initial administration.

The Tenure scale was constructed by identifying 44 items that elicited different response rates from 605 new Target employees who quit voluntarily but would be retained and 654 workers who had been employed at least 3 months with no disciplinary problems.

The Frankness scale items were chosen from an a priori list of statements describing "unlikely virtues" (i.e., statements that describe behavior that is both desirable and nearly impossible to achieve). Twelve such items, showing significantly greater rates of endorsement in the student sample over the job applicant sample, were selected for the final scale.

The Infrequency scale consists of six items that persons are unlikely to endorse unless they are responding randomly or are unable to read English. The average response rate of the items is 4%; only 5% of all subjects gave more than one infrequent response.

Due to item overlap across scales, the final version of the EI contains 97 items. The 97-item EI has been administered subsequently to over 80,000 Target job applicants and to hundreds of job applicants and employees in various other organizational settings. According to PDI promotional literature, a Spanish version of the EI is available, but there are no published reports on this version.

Two forms of the EI are available. On Form A, applicants mark all answers directly on the test booklet. On Form S, applicants respond on a separate answer sheet. The front of the test booklet contains instructions and asks for identification, ethnic background, and gender. The actual items follow on the remaining three pages and are organized into three sections.

Part I contains 69 true-false items worded in the second person. The content of these opinion and attitude statements resembles the content of items on Gough's (1987) Socialization scale and Hogan's (1986) Prudence scale. Part 2 (items 70-83) uses 14 sets of adjective triads from which applicants choose the adjective that best describes themselves. All of the adjectives describe desirable personality traits, but, in each case, one adjective assesses some aspect of conscientiousness, self-control, or honesty. Part 3 (items 84-97) presents 14 multiple-choice "background" items covering previous school and job behaviors and attitudes. Some of these
items inquire directly (but nonoffensively) about past misbehavior and problems in school or work, and some solicit the same kind of information more subtly.

The reading level of all sections of the EI except the adjective triads was estimated with the Flesch reading ease formula. The Flesch score, based upon a count of syllables, words, and sentences, indicated a sixth-grade reading level for the items. The test author notes that some of the adjectives from the triads probably require a slightly higher reading level.

**Practical Applications/Uses**

The EI Employment Inventory was designed for one purpose: to help employers select productive employees and to screen out potentially counterproductive applicants. The test author suggests that the EI will probably be most useful in screening applicants for hourly—especially part-time—positions, but the low pay, lack of benefits, and working conditions of most hourly jobs do not inspire positive employee motivation and loyalty. However, the EI would seem appropriate for any position entailing responsibility (handling cash, merchandise, sensitive information, or others' property; dealing with hazardous materials; looking after the safety and welfare of others). The author claims that research is needed on the usefulness of the EI in occupations in which dependable performance is important, yet it is difficult to think of any such occupations.

Because the EI provides probabilistic results (e.g., Performance scores greater than 65 imply a 2% probability of involuntary termination, a 20% probability of marginal performance, and a 77% probability of 3 months of successful performance), it should not be regarded as a pass/fail test. Rather, EI scores should be considered alongside other relevant information (e.g., background, references, interview) in reaching selection decisions. This point is presented clearly in the administrator’s manual.

Because the scoring key for the EI is proprietary (appearing in neither Paajanen’s administrators manual), one could purchase, at commercial rates, the booklets and necessary computer hardware or software, but the costs could be prohibitive for many academic researchers. Furthermore, research at the item level (internal consistency, factor analysis, etc.) would be impossible. The test author encourages researchers wishing to use the EI in academic studies to contact him personally.

**Technical Aspects**

Reliabilities of the EI scales were estimated in the college sample through Cronbach’s alpha measure of internal consistency and through a 4-week test-retest. Alpha coefficients were .74 for Performance, .64 for Tenure, and .65 for Frankness. (Reliability coefficients for Infrequency are reported but are conceptually unimportant.) These alpha coefficients show acceptable internal consistency, given that the scales were developed empirically over a range of modestly related constructs. (A factor analysis of the first 69 items produced five factors, verifying the multidimensional nature of the EI.)

Because the student sample showed lower scale variance than the employee sample, test-retest reliability coefficients are reported both without and with correction for restriction of range. These test-retest coefficients are as follows: Performance, .62 (.78-.89, corrected); Tenure, .59 (.68-.77, corrected); and Frankness, .67 (.84-.90, corrected).

Students scored a full standard deviation above the applicant sample on the Frankness scale, and a subset of students instructed to complete the inventory “as
honestly as possible" scored higher still. Consequently, a correction for low
Frankness was incorporated into the Performance scale by multiplying the Frank-
ness score by a constant and adding it to the Performance score. "Corrected-
performance scores produced validity coefficients about .05 higher than uncor-
rected scores.

Predictive validity of the Performance scale was demonstrated by correlating scores on this scale with the following ranking of subjects in the employee sample (N = 2,252): 9 = retained for at least 3 months with no significant disciplinary problems; 8 = quit after 3 months but would be hired; 7 = quit before 3 months but would be hired; 6 = fired or laid off but might be rehired; 5 = quit and would not be rehired; 4 = fired and would not be rehired; 3 = fired during probationary period and would not be hired; 2 = fired for minor or serious offenses; 1 = fired for gross misconduct. The (rank order) validity coefficient for the Performance scale with this 9-point criterion was .26 (p < .001). Subjects in category 9 scored almost a full standard deviation above subjects in category 1. Omitting marginal performers (categories 5 and 6) increased the validity coeffi-
cient to .34.

Within the college sample (N = 109). Performance scores correlated signifi-
cantly with all three criteria (all r's < .05). Validity coefficients include an r = .56
with Gough's (1972) conscientiousness test, r = .33 with the Illegal Activity Form, and r = .26 (point-biserial) with defaulting $3.00.

To test the validity of the Tenure scale, subjects in the employee sample who were still employed or who quit voluntarily (n = 1,970) were divided into three
groups and assigned the following rank-order codes: 3 = still employed 3 to 7
months after hire; 2 = quit voluntarily after 3 months but before 7 months; and 1 = quit voluntarily within the first 3 months of hire. Tenure scores correlated r = .23 with this criterion.

Because both counterproductive behavior and early termination can be
accounted for in part by similar personality constructs (e.g., impulsivity), Paajanen expected the Performance and Tenure scales to correlate positively. In-
deed, the scales correlated in the applicant sample (N = 4,669) r = .53. Eliminat-
ing seven overlapping items from either scale produced correlations of r = .46
and .48.

A second test of the validity of the Performance and Tenure scales involved
correlating store averages on these scales with other store variables.

Comparison of Actual Performance to Predicted Performance

553

PDI Employment Inventory

Involving 249 Target stores. Between April and December of 1986, the final version of
the EI was administered to 79,899 applicants, 32,200 of whom were hired. In 207
stores, suggested Performance pass/fail cutoff scores were used for hiring; the
remaining 42 stores continued to use their usual hiring methods without EI in-
formation.

Correlations based on an N of 18,657 between Performance and the same 9-
point ranking used in the original predictive validity study was .21 (corrected for
restriction of range in Performance scores). In the subset of subjects (n = 12,998)
who were still employed or quit voluntarily. Tenure correlated with the same 3-
point criterion used in the original study r = .13. Thus, both empirically con-
structed scales, while demonstrating statistically significant validity coefficients,
show expected shrinkage under cross-validation.

The practical utility of the EI can be examined by comparing store statistics
before and after introduction of the EI. Comparing the 1985 initial research sample
to the 1986 sample, we can see an increase in the percent of employees performing
satisfactory after 3 months (47.9% vs. 24.6%) and a decrease in the percent fired
for minor offenses (.001% vs. .5%), the percent fired for serious rule violations
(.03% vs. 1.1%), and the percent fired for theft (1.3% vs. 2.5%).

Finally, the Performance and Tenure scales have been cross-validated in numer-
ous other concurrent validity studies in other organizational settings. Space con-
straints disallow a detailed description of these studies, but the following are some
representative results. Performance has been found to correlate r = .24 with
overall performance for 72 hard goods retail employees; r = .42 with supervisor
ratings for 30 furniture retail store employees (Tenure correlated r = .55 with this
same criterion); r = .27 with cash variance for 221 bank tellers; r = .35 with supervisor
ratings for 40 Taco Bell assistant managers; and r = .23 with driving performance in 73 metropolitan bus drivers.

The last technical data of concern for the EI involve possible adverse impact
genesis and minorities. Gender differences in both the EI predictor scales and
the performance criteria used in the original study favored females; hence no
scoring adjustment for gender was made. Because females' predictor scores were
5.3 SD higher than males while their criteria scores were 1.7 SD higher, possible
unfairness toward underprediction for females was tested in an independent sam-
ple of 9,570 retail applicants at 90 Sears stores. Examination of residual differences
between predicted and actual performance showed no significant differences be-
tween males and females.

Minority scores approximately .7 SD lower than nonminorities on the Performance
scale, yet show differences in job performance criteria of .3 SD or less.
Minority scores were therefore adjusted by adding 6 points. This correction had
essentially no effect on overall validity coefficients. Analysis of regression resi-
iduals showed no significant differences between ethnic groups. With the race
correction, all protected groups pass the EI recommended cutoff score of 54 at a
rate greater than 89% of the nonminority passing rate, thus showing compliance
with the EEOC four-fifths rule. In the 1986 follow-up study involving 21,519
white, 5,742 black, 3,526 Hispanic, 807 American Indian, and 420 other minority new hires, selection impact ratios were near .99 for all groups, far
exceeding the 80% EEOC rule.
The PDI Employment Inventory is clearly the most carefully developed "personality-based" integrity test available today. Although, in one sense, the EI is only a few years old, it can be regarded as an evolutionary product of outstanding selection research at Personnel Decisions, Inc. and over 40 years of test-construction research at the University of Minnesota. Within a short period of time, the EI has been administered to tens of thousands of applicants and employees and has been cross-validated in over 20 organizations. EI's validity coefficients range from about \( r = 0.20 \) to 0.40. This means that the chance of making an incorrect prediction about a single applicant's productivity can be quite high, but that a group of persons with high scores on the EI will be undoubtedly more productive than a group with low scores. From the standpoint of utility, small productivity savings on each individual can save a large company thousands or even millions of dollars annually.

Given the strengths of the EI, the few negative comments listed below are somewhat picayune. One criticism concerns the uncertain reliability of the EI scales. Given that the EI has been administered to approximately 100,000 job applicants, it is surprising that reliability is estimated with adjusted retest coefficients in a group of 79 college students. Another small problem with the EI is its scoring procedure. Currently, users must mail answer sheets to POI for scoring or can be debated. Adding the points does not affect validity coefficients, but it does produce selection impact ratios far greater than what is required by the EEOC rule. Although the courts almost certainly would not regard the practice as a form of preferential treatment, others may see it that way and wonder if the adjustment is really necessary.

The only real substantive complaints about the EI probably will come from researchers as opposed to commercial users. Because the EI is designed for applied rather than research purposes, practitioners need not consider them. The first problem with the EI from a research standpoint is the invisibility of the scoring procedure, which renders certain types of item-oriented research impossible.

Other researchers will question the low homogeneity of the Performance scale and suggest that factorially derived, unidimensional scales should be used in place of a single scale with low internal consistency. Arguments on the relative merits of factorially complex versus unidimensional scales abroad (Briggs & Cheek, 1986; Hogan, 1986; Johnson, 1988; Wolfe, 1987). For the practitioner, the "best" kind of scale is the one with the highest predictive validity, but only additional research can tell us whether the complex Performance scale will show higher validity coefficients than a combination of narrower, unidimensional scales.

A third problem with the EI for a researcher is that it lacks a bona fide test manual. A good test manual succinctly summarizes the theory, development, technical aspects, and interpretive rules for a test. Currently the EI has an excellent administrator's manual covering administrative procedures and interpretive rules. Unfortunately, one must wade through Paajanen's (1988) dissertation (which is excellent, but long and dense nonetheless) and a series of technical reports (Personnel Decisions, Inc., undated-a, undated-b, undated-c, 1985, 1986, 1988a, 1988b) to learn about the test's theoretical basis, development, and technical aspects. A good personality test manual also provides information on the construct validity of its test by reporting relationships with well-validated measures of personality, interests, temperament, and cognitive traits (e.g., Minnesota Multiphasic Personality Inventory, California Psychological Inventory, Strong-Campbell Interest Inventory, Guilford-Zimmerman Temperament Survey, Wechsler Adult Intelligence Scale). Until this kind of information is gathered and reported, the construct validity of the EI will remain essentially unknown.

A final remark from a researcher's point of view concerns the need for additional research studies as suggested by Sackett et al. (1986). This research would use the EI alongside other integrity tests (a) to determine the amount of shared variance among these tests, especially between personality-based and admissions-type tests; and (b) to determine the relative efficiency of the tests in predicting both global and narrowly defined productivity criteria.

This list contains test citations and suggested additional reading:


References


