POI EMPLOYMENT INVENTORY


Introduction

The POI 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. (POI). 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 5% 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) Personality Attitude Inventory. Forty questions about validity generalization and meta-analysis. Personnel Psychology, 18, 697-709.
426 PDI Employment Inventory

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, 79 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 in the group of 79 students 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 rehired 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 PDt promotional literature, a Spanish version of the EI is available. The EI is composed of three pages and are organized into three sections.

Part 1 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 Hogarth'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 PDI 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 probably will be most useful in screening applicants for hourly—especially part-time—positions, because the low pay, high turnover, benefits of most hourly jobs do not produce 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 unimportant, 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 dissertation nor any PDI technical report), using the EI for research purposes poses some difficulties. This is unfortunate because research on tests like the EI is needed badly (Sackett et al., 1988). If one requires only scale scores from the EI for research, one could purchase, at commercial rates, the booklets and necessary computer hardware, but there are no published reports on the EI by nonacademic 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.

Instructions for administering the EI to job applicants are outlined clearly in the administrator's manual. This excellent manual anticipates a number of questions applicants may raise about the inventory and provides appropriate responses to the questions. Part of the oral instructions directs applicants to written instructions on the cover of the EI test booklet. Applicants should have no difficulty reading or following the instructions unless they cannot read English, they read below the sixth-grade level, or they are handicapped. Average testing time is 15-20 minutes. The Infrequency scale should identify most persons with reading difficulty, yet appropriate alternative provisions for such persons and for people who are blind or otherwise handicapped are not clear. It would seem reasonable to read
items aloud, to explain them if necessary, and to mark answers for handicapped persons, but the validity of this procedure is untested (and probably untestable, given the low base rates both for applications from the handicapped and dismissals for misconduct).

Form A of the EI can be scored in about 2 minutes on site, either by a PDI custom-designed portable computer or by protected IBM-PC software. Form S answer sheets are mailed to PDI for scoring. Hand-scoring templates are not available. To score Form A, the scorer enters, in response to computer screen prompts, the applicant's identifying information and responses to the 97 items. The computer then displays on the screen the applicant's Performance and Tenure scores, which must be hand-copied to the applicant's test booklet.

Infrequence scores of 3 or greater or 20 or more skipped questions automatically flag the protocol as invalid. In either case a message is displayed on the screen indicating why the test cannot be scored. Frankness scores are not reported separately but are incorporated into Performance scores to adjust for exaggerating desirable qualities (see "Technical Aspects" below).

Whether a given Performance or Tenure score is "passing" or "failing" is left to the judgment of the test user, although 54 is the suggested cutoff for Performance. Help for interpreting Performance scores is provided in the form of charts in the administrator's manual indicating the percent of individuals showing satisfactory performance at different score levels. The manual also contains charts showing the probability of staying at least 3 months for different Tenure score levels. Factors to consider when setting relatively high or low cutoff scores are described clearly in the manual.

PDI's computer hardware includes cartridges that will score 50 EIs. When a cartridge is exhausted, it is returned to PDI for recycling. Upon receipt of the cartridge, PDI mails back a list of applicant scores for confirmation. At the end of the year, PDI also provides an annual selection impact analysis report on the proportion of minorities scoring 54 or higher on the Performance scale.

**Technical Aspects**

Reliabilities of the EI scales were estimated in the college sample through Cronbach 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 Infrequence 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 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 Frankness score by a constant and adding it to the Performance score. "Corrected" performance scores produced validity coefficients about .05 higher than uncorrected 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 rehired; 7 = quit before 3 months but would be rehired; 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 rehired; 2 = fired for minor or serious offense; and 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 coefficient to .34.

Within the college sample (N = 109), Performance scores correlated significantly with all three criteria (all ps < .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 = .43 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. Indeed, the scales correlated in the applicant sample (N = 4,609) r = .53. Eliminating 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 comparing store averages on these scales with other store variables. Across 59 stores for which statistics were available, Performance was found to correlate significantly r = .05 or better with loss prevention (r = .27), percent inventory shrinkage (r = .57), termination rate (r = .51), and turnover rate (r = .24).

Tenure correlated significantly with turnover rate (r = .35).

Validity information presented up to this point concerns findings from the 223-item experimental EI. Two things should be noted here. First, the foregoing validity studies do not tell us directly how the final 97-item EI functions apart from the context of the other 126 items. Second, because the Performance and Tenure scales were constructed empirically, validity is "built into" these scales for these two samples. Cross-validation is necessary to determine whether the 97-item final version is valid for additional samples.

Paajanen reports such cross-validational information in a major follow-up study 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 information.

Correlations based on an N of 18,677 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 constructed 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 the introduction of the EI. Comparing the 1985 initial research sample to the 1986 sample, we find an increase in the percent of employees performing satisfactorily 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 (.001% 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 numerous other concurrent studies. Validity studies in other organizations. Space constraints 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 32 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 173 metropolitan bus drivers.

The last technical data of concern for the EI involve possible adverse impact against women 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 .35 SD higher than males while their criteria scores were 1.7 SD higher, no unfairness toward underprediction for females was tested in an independent sample of 9,570 retail applicants at 90 Sears stores. Examination of residual differences between predicted and actual performance showed no significant differences between males and females.

Minorities score 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 residuals showed no pairwise difference between any ethnic groups. With the race correction, all protected groups pass the EI recommended cutoff score of 54 at a rate greater than 80% 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, 186 Oriental, 807 American Indian, and 420 other minority new hires, selection impact ratios were near .99 for all groups, far exceeding the 80% EEOC rule.

Critique

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 20 years 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 = .20 to .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 PDI for scoring or enter applicants' responses into a computer by hand. Scoring would be enhanced by an on-site scan sheet scoring option. Finally, the practice of adding 6 points to minority applicants' scores to ensure compliance with the EEOC four-fifths rule 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 consistencies. Alternatively, one may debate whether a complex Performance scale will show higher validity coefficients than a combination of narrow, 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 exceln-
lent 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. (1988). 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.

References

This list contains text citations and suggested additional reading.


