

## Construction and Validation





# TABLE OF CONTENTS

## **Part 1 — Construction of the Step One Survey™**

<b>Developing the Step One Survey .....</b>	<b>3</b>
The background of honesty testing .....	3
Overt integrity testing.....	3
Personality-based measures.....	4
Validity .....	4
Reliability .....	4
Adverse impact.....	5
Conceptual concerns.....	5
Step One Survey (SOS).....	5
The study .....	6
Study results .....	7
Design of the final instrument .....	7
Discussion.....	9
References .....	11

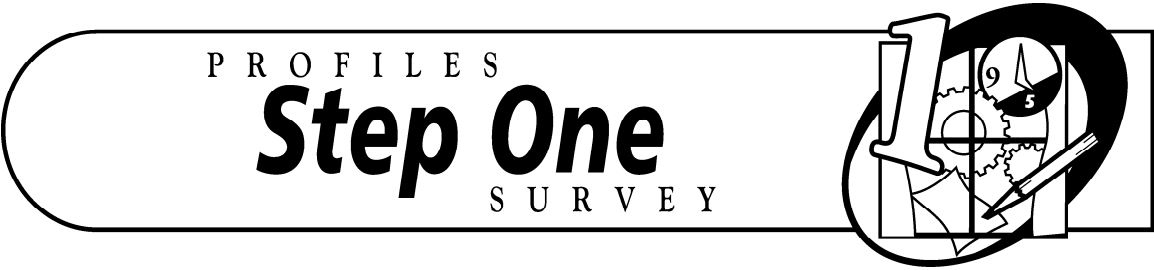
## **Part 2 — Norming and Validation Studies for the Step One Survey™**

List of Tables.....	14
Introduction.....	16
Description of Statistical Methodology.....	17
Results.....	18
Table I — Description of Norming (Employee) Sample .....	19
Table II — Scale Reliabilities.....	19
Table III — Cumulative Percents... Integrity Scale.....	20
Table IV — Cumulative Percents... Substance Abuse Scale .....	21
Table V — Cumulative Percents... Reliability Scale .....	22
Table VI — Cumulative Percents... Work Ethic Scale .....	23
Table VII — Discriminate Validity Scales.....	24
Table VIII — Gender Data .....	25
Table IX — Race Data.....	26
Table X — Age Data .....	27
Summary.....	28

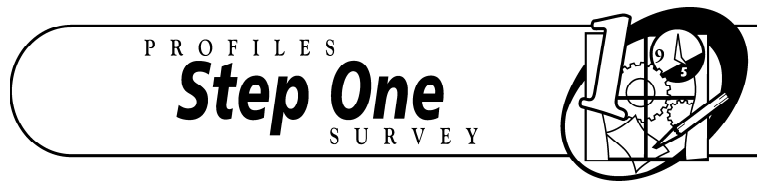
## **Part 2 — Legal Opinion Letter**

Nichols, Wolfe, Stamper, Nally, Fallis & Robertson, Inc.....	29
--	----

**Part 1**  
**CONSTRUCTION OF**



Prepared for Profiles International, Inc., Waco, Texas  
March, 1996



# Developing the Step One Survey

## THE BACKGROUND OF HONESTY TESTING

Paper-and-pencil honesty testing, a multimillion-dollar industry (Tompson, 1981), is designed to curb employee theft, which costs organizations as much as \$40 billion per year (Gorman, 1989). Honesty tests may be administered to job applicants\*. When honesty testing is used prior to employment, employers are attempting to detect potentially dishonest employees.

Originally, these types of instruments were developed as an alternative to the polygraph, the use of which was limited by the Employee Polygraph Protection Act signed into law on June 29, 1988. As an industry, most have been developed outside the mainstream of psychological testing venues by non-psychologists. This limits the information available in standard psychological journals and other forms of third party critiques.

Honesty tests fall into two distinct groups. Sackett et al. (1989) labeled these types as overt integrity tests and personality-based measures.

## OVERT INTEGRITY TESTS

Overt integrity tests are represented in the marketplace by the Stanton, Reid Report, Personnel Selection Inventory, Wilkerson Audit, Phase II and others. As a group, they mix admission and theft-type items targeted to reveal unacceptable attitudes toward employee theft and similar forms of dishonesty in the workplace. With the exception of the Wilkerson Audit, there is no attempt to disguise the purpose of the test. All except the Reid Report include a lie scale to detect faking good. These tests were developed as an alternative to the polygraph examination.

**\* The *Step One Survey* must never be used with people who are currently employed by your organization.**

## **PERSONALITY-BASED MEASURES**

Personality-based measures are represented in the marketplace by the Milby Profile, Personal Outlook Inventory and others. As a group, they usually use standard personality test items first published as scales on the CPI, MMPI, or 16PF mixed with theft items used on overt integrity tests. The rationale is to soften the honesty items and thus disguise them. In general, the composition of the personality-based test items provide information on performance, tenure, reliability, interpersonal cooperation and drug avoidance. Some purport to reveal hostility toward authority, thrill seeking attitudes, conscientiousness, confusion due to vocational identity, social insensitivity, non-conformance, irresponsibility, self-restraint and acceptance of convention. They all include a lie scale to detect faking good.

## **VALIDITY**

Sackett and Harris (1984) reviewed 41 validation studies and grouped the different validation strategies used into five categories: polygraph comparisons, future behavior (predictive validity), theft admissions, shrinkage reduction and comparisons of contrasted groups. Results of the five studies, cited in Sackett and Harris (1984) and Sackett et al. (1989), indicated that honesty tests significantly differentiate between honest and dishonest people.

## **RELIABILITY**

Paper-and-pencil honesty tests are reliable. Their high reliability coefficients compare very well with reliability of other tests in the ability domain (Gatewood & Field, 1987). According to Gatewood and Field (1987), reliabilities for these tests, taking subscales and overall scales into consideration, range from .64 to .94 and mechanical and clerical ability tests have yielded similar reliabilities (.72 to .90). Because personality-based paper-and-pencil honesty tests are broader in focus (Sackett et al., 1989), they have less reliable internal consistency coefficients. It is, therefore, not surprising that their overall reliability (.74) is less than overt paper-and-pencil honest tests (.86).

## **ADVERSE IMPACT**

Adverse impact studies, reviewed by Sackett and Harris (1984) and Sackett et al. (1989), reported no discrimination against protected groups. In fact, some studies reported results in favor of females and blacks. Only age had a significant impact on test performance. Applicants in their teens or early twenties were more likely to score poorly on honesty tests.

## **CONCEPTUAL CONCERNS**

- Comparisons with polygrapher judgments should be dismissed out-of-hand (Sackett et.al., 1984). A criterion in which results are so flawed as to be outlawed by the federal government for pre-employment use cannot serve as a meaningful basis for validation of any instrument.
- Stating that honesty tests can discriminate which applicants will steal from an employer is questionable at best and hard, if not impossible, to prove. The very nature of the exercise is the problem. By screening out those with distorted or unacceptable attitudes, thefts by these individuals that may have occurred do not happen. Therefore, there is no way to prove or disprove the value of administering an honesty instrument.

Predicting how human beings will act in the future is a difficult task. No one can be sure what or why another person thinks in certain patterns. Past behavior may predict future behavior, but the relationship is by no means perfect. Unfortunately, it is the only clue an employer has as to how an individual will act and react on the job. This is the reason that employers check past job references, run background checks and contact personal references before making a hiring decision. It is assumed by the employer that if glowing reports are received from these sources, that the same type of behavior(s) will take place on the job if and when the individual is hired by the company.

## **STEP ONE SURVEY (SOS)**

The objective of the Step One Survey is to add another dimension to the investigative phase of the hiring process with information as to an applicant's attitudes in the areas of integrity, substance abuse, reliability and work ethic. These are ways of thinking which cannot be seen by the interviewer. Just as with job references, background checks and personal references, the only criterion available is past behavior(s). The challenge in developing the instrument was to identify individuals whose past behavior(s) were deemed unaccept-

able by society and employers in the targeted areas. Many other studies have used incarcerated convicts as a standard (Sackett et.al., 1984). The reported flaw in this methodology is that those incarcerated with no chance for release would not take the exercise seriously. Those close to release may try to fake good to enhance their chances for parole.

## **THE STUDY**

The designers of the Step One Survey (SOS) chose the methodology of Contrasting Groups for the validity study. Parolees formulated the group to represent past behaviors unacceptable in society and the workplace. This controlled for the long-term convict and, since participants had already been paroled, it also controlled for those who may try to fake good to get released from prison.

To represent past behaviors acceptable by society and the workplace, a population of employees rated as “ideal” by their supervisors were chosen. Participants had to have been on the job for at least one year and rated as superior in the four targeted areas (integrity, substance abuse, reliability and work ethic).

A questionnaire was created containing items with high content and face validity to be administered to the two populations. It was determined that since past studies using items originally appearing on the CPI, MMPI and 16PF resulted in lower Reliability scores that they would not be used in the SOS. Instead, items were designed asking the opinions of the participants about different issues pertinent to the targeted areas. A lie scale was also added to help detect faking good.

Arrangements were made with the Texas Department of Corrections to administer the new instrument to over 200 convicted criminals as they were released from prison. Age, race, sex and offense information was gathered for EEOC purposes. Their offenses grouped generally as 50% theft, 35% substance abuse, 5% other (kidnapping, murder, weapons offenses, etc.) and 10% unknown. At the same time, arrangements were made to administer the exact same instrument to current employees of a major retailer. This population comprised over 400 employees. Age, race and sex information was gathered for EEOC purposes.

Once all instruments from both populations were received, they were forwarded to Dr. Earl McCallon and Dr. Randy Schumacker at the University of North Texas. They were instructed to complete a Construct and Concurrent Validation Study in tandem with Reliability, Impact and Normalization Studies.

## **STUDY RESULTS**

The completed Validity Study results are in this manual, beginning on page 13. Significant findings included:

- As previously stated, personality-based paper-and-pencil honesty tests are broader in focus (Sackett et al., 1989) and report less reliable internal consistency coefficients of .74, while overt paper-and-pencil honesty tests such as Stanton, Reid Report, Personnel Selection Inventory, Wilkerson Audit, Phase II and others report .86 for internal consistency reliability. The Step One Survey internal consistency reliability estimates calculated for each scale ranged from .84 to .87, the same as overt paper-and-pencil honesty tests.
- The Step One Survey did discriminate between the two groups. Using a stanine scale where 1 is low and 9 is high, over half of the parolees scored 1 on the Integrity scale. On the other three scales, 40% of parolees scored 3 or less.

## **DESIGN OF THE FINAL INSTRUMENT**

To accomplish the purposes of the Step One Survey, it was decided to divide the instrument into two sections. The first section has 45 items and seeks information on self-reported behaviors and admissions, while the second section of 80 items look at behaviorally-based attitudes. Together they combine admission questions, interview questions, personality-like questions and distortion questions. Because the sections compliment each other, it is important for the user to consider the results from both sections in the decision-making process.

### **Section I**

One of the most useful group of items included in overt paper-and-pencil honesty tests are those which ask for the applicant to self-report on past behavior(s). These are referred to as admissions questions. A second concept included on many personality-based honesty tests is the use of paper-and-pencil interview questions. Since the purpose of the Step One Survey is to

help the interviewer in the investigative hiring process, a decision was made to mix these two types of questions into a separate section to precede the psychometric items. Identified as Section I, it consists of 45 items appearing on the final instrument.

## **Examples:**

When did you start work at your current (or last) position?

- |                      |                              |
|----------------------|------------------------------|
| A. 0 - 2 months ago  | D. 1-3 years ago             |
| B. 3 - 5 months ago  | E. Over 3 years ago          |
| C. 6 - 11 months ago | F. This will be my first job |

Have you ever filed a medical claim even though you knew you were not sick or injured?

- A. Yes                      B. Can't remember                      C. No

Since the purpose of the Step One Survey is to furnish the interviewer with more information in order to make a better quality decision, the report gives suggested verbal interview questions based on the answers to the items in Section I.

## **Section II**

Section II contains 80 items. It measures behaviorally-based attitudes towards Integrity, Substance Abuse, Reliability and Work Ethic. Since this is the psychometric portion of the Step One Survey, it was subjected to validation/reliability studies by the University of North Texas.

Examples of these attitudes include:

- Sloppy company security causes some people to steal.
- A person doesn't have a drug problem if all he/she does is smoke marijuana on weekends.
- If you have a good excuse, it's okay not to show up to work.
- Loyalty to a company is a thing of the past.

## **Distortion Scale**

Because the final instrument was designed to be used with job applicants, it was important to include some way to determine if the applicant was revealing his/her true feelings and not what he/she considered the correct answer. Therefore, a scale labeled Distortion was incorporated throughout the instrument. This scale contains items for which the correct answers are already known. The applicant must give the correct pre-determined response to the majority of items on this scale for the results to be considered an indication of the applicant's true feelings in the selected areas being measured. This distortion scale also looks for inconsistencies in the response pattern.

The distortion scale is designed to reflect obvious lies and/or inconsistencies in the responses on Section II of the survey. If the score obtained on this scale suggests a distortion problem on Section II, the user is asked to weigh carefully the accuracy of the responses given in Section I. While it would be unusual to expect an exaggeration of negative admissions on Section I, the reverse is not true.

## **DISCUSSION**

There may be a question as to why Parolees were used as one of the contrasting groups because they are a subgroup that may not really represent applicants seen by employers. Another question might be that all the Step One Survey is doing is separating out this subgroup from "normal" applicants, not discriminating as to who will steal from the employer if hired. Some thoughts on these questions follow:

- Parolees were chosen to represent a group of individuals whose past behavior(s) indicate patterns of thinking unacceptable by society or in the workplace. Even those asking the questions would agree that in all probability there are numerous individuals in our society who have not been convicted or gone to prison even though they think in the same anti-societal patterns. Unfortunately, there is no way to discriminate these applicants from those with so-called "normal" attitudes since the interviewer cannot tell which applicants think in these distorted patterns. Therefore, using a group known to have these types of attitudes establishes a needed base line for comparison. Using the Parolee's group scores as a base line does not keep an individual Parolee who does not think in these distorted patterns from obtaining a high score on any of the Step One Scales. In fact, some of the Parolees did attain acceptable scores on all the scales.
- It is not the intended purpose of the Step One Survey to make an absolute judgement about

## ***Construction and Validation of the Step One Survey™***

applicants who will steal if hired by the user-company. None of these types of instruments can make that claim. In fact, no one can be totally accurate about what another person will do in any given situation, because no one can tell what another person is truly feeling or thinking at any given time. The purpose of SOS is to help the interviewer by furnishing valuable information as to how an applicant thinks in four targeted areas. SOS results are only a part of the hiring decision. There are no cut-off scores, no pass-fail scores. The final decision, after reviewing SOS results, remains with the interviewer.

It is suggested that the user weigh the Step One Survey results as only one-third (33%) of the decision process. The balance is suggested to be evenly divided between history (resume, references, background checks, etc.) and interview results. This combination gives the user a balanced view of the applicant.

## References

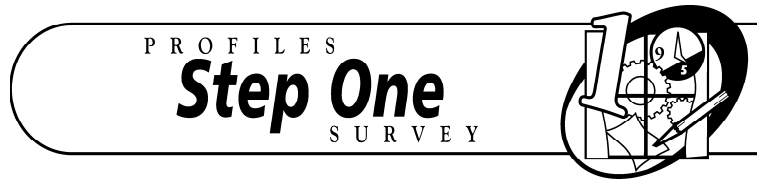
Gatewood, R. D., & Feild, H. S. (1987). **Human Resource Selection**. New York: The Dryden Press.

Gorman, Christine (1989). "*Honestly, can we trust you?*" **Time**, p 44.

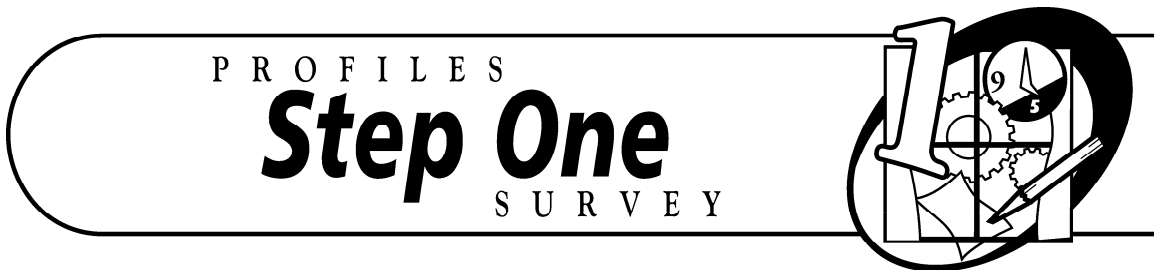
Sackett, P. R., Burriss L. R., & Callahan, C. (1989). "*Integrity testing for personnel selection: an update.*" **Personal Psychology**, **42**, 491-529.

Sackett, P. R., & Harris, M. M. (1984). "*Honesty testing for personnel selection: A review and critique.*" **Personal Psychology**, **37**, 221-245.

Tompson, S. (1981). "*More employers attempt to catch a thief.*" **The Wall Street Journal**, August 3, 1995.



**Part 2**  
**NORMING AND**  
**VALIDATION STUDIES**  
**FOR**



**Dr. Earl McCallon, Professor**  
**University of North Texas**

**Dr. Randall Schumacker, Associate Professor**  
**University of North Texas**

Prepared for Profiles International, Inc., Waco, Texas  
March, 1996

## **LIST OF TABLES**

- TABLE I:** Composition of Norming (Employee) Sample with Regard to Age, Gender and Race
- TABLE II:** Means, Standard Deviations and Alpha (Reliability) Coefficients for Each Scale on the **STEP ONE Survey™** Instrument
- TABLE III:** Stanine Scores Achieved by Norming (Employee) and Parolee Groups on the Integrity Scale
- TABLE IV:** Stanine Scores Achieved by Norming (Employee) and Parolee Groups on the Substance Abuse Scale
- TABLE V:** Stanine Scores Achieved by Norming (Employee) and Parolee Groups on the Reliability Scale
- TABLE VI:** Stanine Scores Achieved by Norming (Employee) and Parolee Groups on the Work Ethic Scale
- TABLE VII:** Means, Standard Deviations and t-Test Values for Parolees and Norming (Employee) Sample Scores on the **STEP ONE Survey™** Instrument
- TABLE VIII:** Means, Standard Deviations and t-Test Values by Gender for the Norming (Employee) Sample
- TABLE IX:** Means, Standard Deviations and Anova F-Ratios by Race for the Norming (Employee) Sample
- TABLE X:** Means, Standard Deviations and Anova F-Ratios by Age Categories for the Norming (Employee) Sample

## **The Distortion Scale Score**

The Distortion score applies **ONLY** to Section Two. However, if the distortion scale suggests distortion on Section Two, then some doubt may be cast upon the openness of the applicant in their responses on Section One.

The Distortion score refers to the **reliability of the results, not the honesty of the individual**. A low score on this scale suggests that for some reason the applicant may have distorted their responses in the “look good” direction. This could possibly happen because of an attempt to portray a picture of how they would like to be seen, rather than an accurate picture of how they are.

With a Distortion score of 7, 8 or 9, no obvious attempts to “look good” were detected. With Distortion scores in the 4 to 6 range, the results may reflect a somewhat “polished” version of reality. When the Distortion score reaches the lower end of the scale, 1, 2 or 3, then the **STEP ONE Survey™** results should not be a part of the decision-making process for that individual.

The suggested interview questions found in the **STEP ONE Survey™** report should be pursued in all cases where an individual continues to be considered for a position. When the applicant has openly admitted to a behavior, the discussion of that behavior may prove fruitful to the hiring process.

## **INTRODUCTION**

This report presents the results of a norming and validation study for the **STEP ONE Survey™** instrument. The studies described in this report are follow-up studies to a series of studies conducted earlier that determined the factorial structure of the **STEP ONE Survey™** instrument and presented selected validity and impact analyses. This report used the original data collected in the fall of 1995. These data consisted of 411 retail store employees (referred to as norming group/sample in this study) and 210 recent prison parolees.

The current study sought to determine the best measurement items on the four scales underlying the **STEP ONE Survey™** instrument, calculate norms for the retail store employees, and analyze the performance of the parolee groups on these norms. Since any reduction in the number of items on a scale would be expected to impact the reliability and validity estimates of the scales established by the earlier study, new reliability estimates were calculated on the reduced set of items for each scale. New discriminate validity studies were also conducted with respect to gender, race and age. These results are presented in this report.

This study was conducted using the strict standards and guidelines established by the American Psychological Association (APA) for Test Development.

## **DESCRIPTION OF STATISTICAL METHODOLOGY**

Factor loadings and communality estimates from the original factor analysis study using the principal components estimation method with varimax rotation were used to determine items on each scale that best represented that particular factor. Once these items had been determined, Cronbach Alpha internal consistency coefficients were computed to assess the reliability of each scale.

The second series of analyses consisted of impact studies for the age, race, and gender variables. The age and race variables were analyzed using one-way analysis of variance and Scheffe post-hoc procedures. Gender differences were studied using the independent t-test. These procedures were performed on each scale on the **STEP ONE Survey™** instrument.

Once the above analyses had been performed, norms were calculated using the retail employee subjects. Raw scores were converted to normalized z-scores and equivalent starting scores were determined. The parolee group's performance was then determined and compared to the norming (employee) sample. It was hypothesized that the relative performance of the parolee group would be lower than the norming (employee) group.

























