TESTING

Theoretical and Applied Perspectives

Edited by
RONNA F. DILLON and
JAMES W. PELLEGRINO

PRAEGER

New York
Westport, Connecticut
London

9 Situationist Charges versus Personologist Defenses and the Issue of Skills

Luc Bovens
Arnold Böhrer

In this chapter we intend to discuss a few theoretical issues in personality theory and in the methodology of personality measurement. On the background of this discussion we will then present some of the personality tests that were developed (at least partly) within the Section for Psychological Research in the Center for Recruitment and Selection of the Belgian Army.[1]

SITUATIONISM VERSUS PERSONALITY THEORY

Situationist Charges and Personologist Defenses

The first theoretical issue that we are planning to consider is whether the very concept of personality (1) can play a useful role in contemporary psychology or (2) is merely an aggravating legacy from *folkpsychology* (that

We are grateful to the numerous people who have in some way contributed to the research on the free-format self-description method (VZM). We would especially like to mention P. De Boeck, R. Biesmans, S. Van Den Broucke, and the Laboratory for Personality Research in the Dept. of Psychology in the K. U. Leuven. For research on the ROMAT and the RS1C we respectively thank Ph. Schotte and P. Vermeulen.

is, the common art of explaining or predicting actions, emotions, etc., which people generally practice in living their lives).[2] Here is how the debate runs. Let's start off with a caricature of the latter position, which we promise to label "extreme situationism." The extreme situationist claims that all psychological explanation that makes reference to persistent personality traits:

(b_1) is merely ad hoc and

(b_2) cannot be but unsuccessful, since there simply is no constancy in a person's behavior when one is confronted with differing environments.

Let us at first consider objection (b_1). Molière ridiculed the science of his age in a play which stages a medical student proclaiming solemnly that opium puts a person to sleep due to its *virtus dormitiva* or somniferous power (Molière, 1947, p. 196). Now the extreme situationist claims that this is precisely the kind of situation that the personality theorist is in, say, in explaining hysteric outbursts in reference to a neurotic character and that all personologist theorizing actually has the same ad hoc character to it.

Consider now objection (b_2). In trying to explain or predict some action, folkpsychology does pay due attention to person-variables ("What sort of person are we dealing with?"). Now extreme situationists claim that this folkpsychological mode of explanation or prediction should not be adopted in science, since all action is, as a matter of fact, determined only by situation-variables. Across sufficiently differing environments there simply is not constancy in a person's behavior, and consequently, person-variables lack all explanatory or predictive power. Thus, if the situationist objections hold, the personality theorist would even be worse off than Molière's scientist. Reference to a somniferous power is not very informative, though it is reasonable to assume that opium will have the same power tomorrow as today. The personality theorist, on the other hand, cannot even assume that the ad hoc features which he calls in have any persistence over time.

Now, what kind of response does the personality theorist offer to these extreme situationist charges in defense of his or her project? The personality theorist believes that personality traits are a necessary component in psychological theorizing. The personality theorist, in granting that the situationist objections may more or less hold against folkpsychological traits, believes that (a_1) situation-variables, taken by themselves, are insufficient to explain or predict actions and (a_2) that the situationist objections can be overcome if we are sufficiently careful in selecting the set of personality traits that are to play a role in psychological theory.

What sort of traits then are to play a role in proper psychology? Within personality theory itself, there are widely divergent opinions concerning what should count as an adequate personality feature in psychological theory. Consider the set of categories in our common practice of personality trait ascription, that is, the set containing descriptive terms like "honest,"

"friendly," "outgoing," etc. Now, this set of folkpsychological terms is apparently inadequate for doing proper psychology. But how should we adjust this set?

There are two such psychological projects which we would like to discuss within this chapter, namely, the construction, on the one hand, of a set of situation-specific traits and, on the other hand, of a set of five personality factors. The first project specifically aims at insuring the *constancy* of personality traits, though we believe the price for this is that a psychological explanation which makes reference to such situation-specific traits even has more of an ad hoc character than folkpsychological explanations. Conversely, the second project reduces the ad hoc character of psychological explanations, though the price for this may well be that such factor has even less constancy than folkpsychological features.

Let us now take a close look at both projects. The first project is connected with the name of Lawrence A. Pervin. Pervin (1976, 1977) asks his subjects to:

1. list some significant situation in their current lives;
2. describe these situations;
3. describe how they feel about each situation;
4. describe how they behaved in each situation.

He then isolates four sets of factors within the set of situations for each particular individual. In the first set similar descriptions load on the respective factors, in the second set similar feelings, in the third set similar behaviors, and in the fourth set similar combinations of descriptions, feelings, and behavior. Now this analysis yields four lists of situation-specific trait decriptions for each person. These respective lists contain the following types of statements.

List 1: Person X is the kind of person who typically *describes* y-related situations as z_1.

List 2: Person X is the kind of person who typically *feels* z_2 about y-related situations.

List 3: Person X is the kind of person who typically *reacts* z_3 to y-related situations.

List 4: Person X is the kind of person who typically, for y-related situations, describes them as z_1, feels z_2 about them, *and* reacts in a z_3 way to them.

The set of factors y includes things like school, home, friends, etc.; z_1s are adjectives descriptive of situations, such as comfortable, frustrating, demanding, etc.; z_2s are adjectives denoting emotions, like happy, nervous, scared, etc.; z_3s are adjectives denoting types of behaviors, like responsible, friendly, irascible, etc.

Now this set of lists will yield a very detailed and structured description of what kind of person X is. We can also generalize concerning the rate of

occurrence of certain descriptions, feelings, and behaviors, or concerning particular interrelations between descriptions, feelings, and behaviors. These empirical data show that for a particular person some traits are highly general, that is, hold across a broad set of situation-bound, that is, hold only for a small set of situations. Such a situation-specific trait description, that is, a trait description that includes reference to the range of situations to which the trait applies, is Pervin's alternative to folkpsychological trait description. And this mode of trait description, Pervin argues, can (at least partly) avoid the situationist objection that traits have no constancy. Now this indeed seems to be a reasonable claim. Most people certainly do adjust their styles to different types of situations. Consequently, Pervin's situation-specific trait descriptions will do better on the issue of constancy than our much less situation-specific folkpsychological trait descriptions.

On the other hand, we do have serious doubts about the explanatory value of such traits. It seems to us that in explaining an action, the more situation-specific the explanatory trait becomes, the more we are dissatisfied with its ad hoc character. Here is an example: We are interested to know why Mary blew up at her coach during soccer practice when the coach told Mary she expected more commitment from her. Now following Pervin's model we may say that Mary is the kind of person who reacts irritably (z_3) in sports (y) when she feels personally attacked (z_2) in a threatening situation (z_1). But is this a genuinely satisfactory explanation? It seems to us that what we would really want to know is *why* Mary is precisely such a kind of person. Now answering this question forces us to use more general, and thus less situation-specific trait descriptions. We may want to say that Mary is a very quarrelsome personality, who would construe any situation as being threatening just in order to pick a fight. Or maybe Mary is just in general a very sensitive character, who is very much affected by remarks concerning her performance. Or maybe Mary is very competitive, such that her whole life turns around sports: consequently she takes such comments by her coach very personally. Now we think it is these sorts of explanations which would genuinely satisfy us. But then of course, with more general personality-traits, the problem of constancy again becomes more acute.

Let us now turn to the 5-personality-factor theory. Tupes and Christal (1961) presented their subjects with a checklist with 35 pairs of bipolar trait-names (silent vs. talkative, depressed vs. cheerful etc.) This set of pairs they considered to be representative of the full set of trait-names in the English language. The subjects in their study were asked to check the adjectives in each pair which they considered to be most descriptive of their personality. A factor analysis yielded five *factors* or sets of intercorrelated pairs of adjectives, which were labeled "surgency," "agreeableness," "independability," "emotional stability," and "culture." Now much research has been done on fine-tuning and redefining these categories, as well as on identifying the same set of factors in data from other measurement-techniques

(questionnaires, peer-ratings) (for example, McCrae et al., 1985a, 1985b, 1986, 1987). McCrae and Costa claim that these five factors (or actually, a close variant) provide for the basic structure of personality and that any model of personality *must* include at least these five factors. Whether it should include more depends on the scope one wants to assign to the concept of personality, for example, the five-factor model does not include cognitive abilities or elements of self-concept like body image or social identity (McCrae & Costa, 1986, p. 22). Now this project reduces the set of folkpsychological concepts to an extremely small set. What are the consequences if we decide to do proper psychology with this particular set? On the negative side, it seems to me that this project is more vulnerable to situationist charges of lack of constancy. Consider a person who scores on average on agreeableness. The adjective pairs "suspicious vs. trustful" and "cool, aloof vs. attentative to people" load on the factor of agreeableness. But now imagine this person is very agreeable when it comes to being sensitive to other persons' needs, while not at all agreeable when it comes to trusting other people if it is sensible to do so. In this particular case, the ascription of the feature "being agreeable" in five-personality-factor theory would display less constancy than the folkpsychological ascriptions "being suspicious" or "being attentive to people." On the positive side, it seems that if it were the case that we could construct models in which the dependent variable is some type of action and the independent variables includes a combination of scores on the five-factor scale (plus a set of situation-variables), this would provide for an explanation which supersedes the charges of the ad hoc character of explanations involving personality-features. Situationist charges of ad hoc-ness would become obsolete vis-à-vis explanations invoking general statements of the form "If some person scoring such and such on surgency, agreeableness . . . is placed in ____ circumstances, she is likely to do ____."

Now, what can be concluded from this discussion? We have argued that extending the set of folkpsychological trait names by making them situation-specific may alleviate the problem of constancy while intensifying the problem of the ad hoc character of the explanations thus obtained. On the other hand, reducing the set of folkpsychological trait names to five factors may alleviate the problem of the ad hoc character of the explanations thus obtained, while intensifying the problem of constancy. Now it may thus well be the case that our common folkpsychological trait names provide for an optimal balance between both alternatives. The wisest route to take for proper psychology may thus be to work with our common folkpsychological trait names or close variants in constructing explanatory models. And this, I believe, is what most often happens in actual psychological theorizing anyway.

An Application in Personality Testing: The VZM

The concrete problem we were facing was the development of a personality test for the purpose of selection within a military setting. Now what can we learn from our theoretical reflections as to what features are desirable for a

personality test with this particular purpose? In the first place we are interested in a few traits that our folkpsychological intuitions tell us might be good predictors for a successful military career, namely, *leadership* and *creativity*. Furthermore, we are also interested in sketching an overall picture of the test taker's personality. For this purpose we decided it would be desirable to collect scores on a more recent variant of Tupes and Christal's five-personality factors, involving *extroversion, agreeableness, conscientiousness, neuroticism,* and *general culture* (McCrae & Costa 1985a, 1985b; McCrae et al., 1986, 1987). But if we decide to adopt this five-factor model, we must find some way to (more or less) bypass the situationist charges of no constancy. Pervin's research provides for empirical evidence that, while some traits are highly situation-bound, other traits are quite persistent across situations for particular persons. Now it is precisely this last set of traits that we are interested in, in sketching an overall picture of our test takers. But how can we get at precisely these traits? We assumed that, if a person would have a chance to describe herself by means of a list of adjectives, she would be most struck precisely by these character features that affect her life across differing environments. Now, if this assumption holds, free self-description would thus be an appropiate technique to get at more or less constant personality traits.[3] A particular problem of a self-descriptive method in a context of selection is that the test takers will attempt to present an overly positive picture of themselves. In order to check for this tendency we have added one control-dimension, *social desirability*. In order to make our results interpersonally comparable it would be desirable to assign standardized scores on all these dimensions. But now how can we translate such widely diverging lists of self-descriptions into scores on the particular dimensions which we are interested in, namely, *leadership*, *creativity*, the five-personality factors, and *social desirability*? In order to solve this problem we have worked out a test procedure, the VZM or *V*rije *Z*elfbeschrijvings*m*ethode (the Free-Format Self-Description Method) which closely resembles Potkay and Allen's Adjective Generation Test (AGT) (1973).

The subjects in our test are asked to describe their personality by means of ten adjectives. These adjectives are then looked up in a scoring list. The scoring list contains a large set of "expressions," that is, nouns, adjectives, and descriptive phrases, which might possibly describe a personality feature, with matching scores assigned by groups of judges on each of the eight dimensions. The subject's overall raw score per dimension is then calculated by adding the respective scores for each adjective. These raw scores are then converted and standardized into an 11-point scale with fixed mean and fixed standard deviation. In a last step, these standardized scores are plotted on a graph, which provides for a personality profile of the person in question.

Let us now consider some details. The expressions in this scoring list were compiled from a set of responses from some pilot studies involving more

than 3,000 subjects. The scores were obtained by having groups of ten judges rate some hypothetical person who would adopt the expression in question to describe himself, on *social desirability*, on the five factors, on *leadership*, and on *creativity*. If the judge thought the hypothetical person would (not) have the feature in question to a large degree (at all), a score of +2 (−2) was assigned to the particular expression. Intermediate scores were available of +1 and −1, and a score of zero was assigned if the expression did not allow for a judgment on the feature in question. Scores were added over the ten judges and an overall score (between +20 and −20) was assigned for each expression per dimension. Here is an example:

	Ambitious	**Persistent**
Social Desirability	7	7
Extroversion	5	3
Agreeableness	-2	-1
Conscientiousness	3	9
Neuroticism	-2	-5
General Culture	7	3
Leadership	13	10
Creativity	1	2

To improve the uniformity of scoring in the judging task we attached some descriptive phrases to each dimension (apart from *social desirability*). For example, for *creativity* we added the following specifications: "inventive, imaginative, breaks through old patterns of thought, approaches issues from different angles and comes up with new ideas."

The subjects are encouraged to write down ten adjectives. Some subjects, however, choose expressions which are not adjectives and some hand in lists of less than ten responses. In the scoring list we have included expressions other than adjectives to allow for the former possibility. Furthermore, if a response does not occur in our list, the test examiner may replace this response with a synonym which does occur in the list. (We know from current research that only 10 percent of the responses was not included in the Dutch list.) If no synonym can be found, the response may simply be dropped from the list. (Research has shown that dealing with the problem of responses which do not occur in the list of expressions by either dropping all such responses, searching for *all* such responses, or setting a group of ten judges to work at rating the particular responses did not yield significantly different results. We have thus opted for a method which combines the ease of the first method and the frugality of the second, while avoiding the complexity of the last method.) If the adjusted list contains a total of five or more responses, overall scores for the subject in question can meaningfully be calculated. Scores should then be adjusted to a basis of ten responses such that they become comparable between subjects.

Since the raw overall scores taken by themselves do not allow for a meaningful interpretation, we have worked out a standardization procedure. This procedure allows for a conversion of raw scores to C-scores, that is, standardized scores on an 11-point scale ranging from zero to ten, the mean score is five and standard deviation is two. Previous test results have made it possible to construct tables for score conversion for particular populations (for example, for candidate reserve officers).

Let us now consider some of the reliability and validity studies on our test method. We have run two *test-retest* reliability studies on candidate officers. The time lapse between testing and retesting was respectively between six and 12 months and two days. The former study yielded a test-retest correlation of .50, the latter of .77. We did a study on the *scoring reliability* by having five persons score the responses of 100 candidate professional officers. (Remember that in choosing synonyms the scorers have to make some subjective judgements.) If the poorest scorer was deleted from the correlation-matrix, all inter-scorer correlations per dimension were higher than .90. The *score reliability* was checked for by having two groups of ten judges score 24 randomly chosen expressions. The correlations were .95 for *extroversion*, .95 for *agreeableness*, .89 for *conscientiousness*, .84 for *neuroticism*, and .85 for *general culture* (yielding .90 on average).

The *congruent validity* of the VZM was checked for in two studies. In a first study with 77 senior high school students we calculated correlations with the five-personality factor test (Elshout & Akkerman, 1975). Correlations per dimension were .71 for *extroversion*, .54 for *agreeableness*, .60 for *conscientiousness*, .55 for *neuroticism*, and .55 for *general culture*. In a second study (Böhrer & Van Den Broeck, 1986) with 63 candidate reserve officers, we calculated correlations between the eight VZM- dimensions and a set of questionnaires in our selection procedure including the PMT of Hermans (1976) (measuring achievement motivation (P), and debilitating (F-) and facilitating (F+) fear of failure); the social anxiety scale of Willems, Tuender-De Haan, and Defares (1973); and the ABCA questionnaire of Böhrer (1980) (measuring social anxiety and self-confidence). The correlations are presented in Table 9.1. We hope these results can convince our readers that the VZM, as a new method of personality research, can stand up to questionnaires as far as reliability and validity goes. Aside from the theoretical justification of the VZM, which we have set up earlier in this chapter, some more practical considerations should also be mentioned in defense of our new test method: The VZM is not as time-consuming as questionnaires, as well as more attractive for the test takers, since they feel they can express themselves freely using their own words and there is no time pressure.

SKILLS: THEORY AND TEST DEVELOPMENT

Can Skills Be Reduced to Personality Features Plus Aspects of Intelligence?

I will now turn to the second theoretical issue. There are all kinds of folkpsychological features which we commonly ascribe to persons, for

Table 9.1
Correlations between VZM-Dimensions and Some Questionnaires

		VZM-Dimensions						
	SD	EXT	AGR	CON	NEU	GC	LEA	CRE
PMT P	.36	.24	.07	.32	-.17	.29	.34	-.02
F-	-.48	-.08	-.11	-.26	.41	-.33	-.41	-.04
F+	.21	-.01	.09	.21	-.14	.09	.23	-.03
Social Anxiety Scale	-.36	-.32	-.09	-.16	.32	-.32	-.45	-.25
ABCA Social Anxiety	-.42	-.58	-.30	-.24	.16	-.22	-.44	-.28
Self-Confidence	.37	.09	.01	.35	-.23	.25	.37	-.08

example, "Alice wants to become a medical doctor," "Joe felt lonesome yesterday night," "Mary is extremely bright at math," etc. Psychologists tend to cut up this set of features into two subsets, namely, a subset of cognitive features and a subset of noncognitive or personality features. Within the former set they distiguish between beliefs on the one hand, and intelligence on the other hand. The latter set contains a mixture of conative features, reflecting our values and aspirations, as well as *affective* features, depicting our emotional lives.

Now here is a problem concerning this dichotomy between cognitive versus noncognitive features. There exists a large set of mental ascriptions that can neither be classified as *pure* cognitive, nor as *pure* noncognitive features. I am particularly interested in the ascription of *skills*, say leadership skills, or the skill of learning a second language, etc. Now many such terms, which stand for skills, are *mixed* terms, that is, they neither denote *pure* cognitive, nor *pure* noncognitive features. Furthermore, for at least some subset of such mental ascriptions, it does not seem possible within folkpsychology to reduce such ascriptions nontrivially[4] to some mixed set of terms denoting either pure cognitive or pure noncognitive features. Many of our common terms denoting skills are irreducible in our folkpsychological language, that is, we simply lack the common vocabulary to replace names of skills by mixed sets of pure terms. In other words, for some names of skills, it is the case that the mental ascription "having such and such skill" (say, the skill for learning a second language) is *not* just reducible to a particular set of *common* terms denoting personality features (say "being sociable") and *common* terms denoting cognitive features (say, "having a good memory"). We want to say that having a skill for learning a second language is something more than just the sum of being sociable and having a good memory, and furthermore, whatever pure *common* terms we add to this sum, it will never equal the skill in question. In folkpsychology, there

are at least some skills which are what they are and nothing else!

Now within proper psychology this state of affairs may prompt three kinds of responses:

1. We may decide to fine-tune our folkpsychological set of *pure* cognitive and (or) noncognitive vocabulary with the aim of making a reduction feasible within proper psychology. I propose to label this position "*narrow reductionism.*"
2. We may decide to give up the reductionist project and include folkpsychological ascriptions of skills as irreducible terms in our psychological vocabulary. Let us call this the *antireductionist* position.
3. We may decide to construct a limited psychological set of special mixed terms, that is, terms that are in some sense elementary as well as general, though do include reference to *both* cognitive and noncognitive features. Such terms are said to denote types of *cognitive style.* The aim of constructing such set is to give the reductionist program more leeway: it is claimed that our folkpsychological ascriptions including pure cognitive terms, pure noncognitive terms, and (or) mixed terms of cognitive style. This position I propose to call "*broad reductionism.*"[5]

Now, the assessment of (potential or actual) skills is a major concern in psychometric research, considering its practical applications for screening job applicants and for career-counseling. Now I believe that each of these theoretical stands on the issue of reductionism has its correlate in measurement techniques for the assessment of skills. In our own research we have developed two tests which respectively match the antireductionist and the broad-reductionist project. At first, however, I will relate the narrow-reductionist program to a classical technique in measurement theory, namely, empirical criterion keying.

The Narrow-Reductionist Project and Empirical Criterion Keying

Consider a (sympathetic) layperson's views on IQ-tests or personality tests in a context of selection or career-counseling. The layperson believes that the psychologist knows how to translate, say, "the skill of being a good sales manager" into a set of personality features, such as, "being meticulous" and "being assertive" and attributes of intelligence, "being proficient in mathematical reasoning." Furthermore the psychologist is thought to know how to translate raw test results into claims concerning the degree to which the testtaker can be assigned such personality features or attributes of intelligence. It is this *double-translation job* that makes it possible for the psychologist to uncover potential skills by means of apparently trivial test questions.

Now the layperson's views on measurement theory hinge on the belief (1) that skills can be translated into sets of common personality features and

attributes of intelligence and (2) that sets of raw test results can be translated into claims concerning common personality features and attributes of intelligence. Counter to the first belief, I have already indicated that our folkpsychological vocabulary is too impoverished to allow for this sort of crude reductionism. Counter to the second belief, our concern about the content-validity of measurement techniques comes to show that this translation job is not always fully translucent.

Empirical criterion keying (cf. Anastasi, 1961, pp. 528-541 for empirical criterion keying on interest tests) cleverly avoids dependency on both precarious beliefs. How so? Some groups will score consistently higher on particular IQ questions and respond to personality tests in idiosyncratic ways. Groups of successful people within certain professions are presented with personality tests and/or IQ tests. For each such profession the psychologist then tries to identify particular scoring patterns. Within a context of selection, candidates for some jobs are then chosen because their particular scoring pattern most closely resembles the typical scoring pattern of some sample of persons successfully performing the job in question. Within a context of career-counseling, the future student's scoring pattern is compared to typical scoring patters for a number of jobs in order to locate the branch of study which would suit that student most.

The skill for performing certain professions is thus reduced to the fine-tuned language of scoring patterns in IQ tests and/or personality tests. In folkpsychology, it was impossible to express skills in terms of our common vocabulary denoting aspects of intelligence and personality features. Now empirical criterion keying (vis-à-vis skills) is the project in proper psychology which tries to reduce (potential) skills to a combination of aspects of intelligence and personality features, though described on terms of the sophisticated vocabulary of scoring patterns in psychological tests. In this project, the need for translating test results *or* skills in terms of folkpsychological terms denoting aspects of intelligence or personality features becomes obsolete.

It is clear that empirical criterion keying is an attractive project. There is the simplicity of analyzing skills only in terms of aspects of intelligence and personality features and the clever avoidance of the precarious belief in the possibility of a satisfactory execution of the double-translation job. But is this attractive project feasible? What price do we need to pay for such elegance? Here are a few conceptual considerations which made us decide not to choose for the narrow reductionist route in test development. First, there are certain personality features and/or aspects of intelligence which are causally, though not essentially related to the successful performance of a particular job. Personality features and aspects of intelligence which are essentially related to the successful performance of some jobs are—on the narrow-reductionist project—identical to the skill in question. Personality features and aspects of intelligence that are causally, though not essentially related to the successful performance of a job, are merely by-products of

the skill in question. Empirical criterion keying does not allow us to distinguish between aspects of intelligence and personality features that are essentially versus nonessentially related to the successful performance of some job, or in other words, it cannot distinguish between the skill and its by-products. Please allow us to invoke a common and probably incorrect stereotype to illustrate this point. Librarians are often considered to be dull personalities. Now assume—counter to our honest expectations—that this personality feature would show up through empirical criterion keying. Now would we then say dullness is part of the skill of being a good librarian? I do not think so. Rather, we would say that the methodical work librarians do is casually efficacious in bringing about their dullness, but dullness is therefore not an essential feature of being a good librarian.

Second, it may well be the case that, for some jobs, the aspects of intellligence and personality features that are essentially related to "having the potential skill for some job" and "having the actual skill for some job" are very different or even incompatible. Consider some artistic discipline, for example, ballet. Most likely, creativity is an essential feature of the actual skill of a talented ballerina. But now this same creativity may be counterproductive in a beginning class for ballet. What we need the most in this context is an obedient and rigid performance of basic exercises. The personality feature that is essentially attached to the potential skill for ballet is thus very different from the personality feature attached to the actual skill for ballet. A talented ballerina does not start off as a creative dancer, but *acquires* this very creativity through rigidity. Now empirical criterion keying can solely tell us about the personality features and aspects of intelligence related to *actual skills*. For those personality features and aspects of intelligence for which it is essential that they are acquired *during* the job-training, empirical criterion keying can solely yield a misleading test procedure for assessing potential skills.

Finally, if skills *are* actually more than a combination of personality features and aspects of intelligence, empirical criterion keying on personality tests and IQ tests may leave some feature that is genuinely essential to having the skill in question, unassessed.

For all these reasons we have opted for steering our projects in test development away from the narrow-reductionist route.[6] We have worked out two tests that respectively match the antireductionist and the moderate-reductionist route. Both tests will now be discussed.

The Antireductionist Program and the ROMAT

In this section we will consider the methodological correlate of the antireductionist position. The antireductionist measurement theorist claims that any set of information obtained from cognitive tests and from personality tests (whether expressed in folkpsychological terms or in terms of scoring patterns) will be insufficient for properly assessing particular skills.

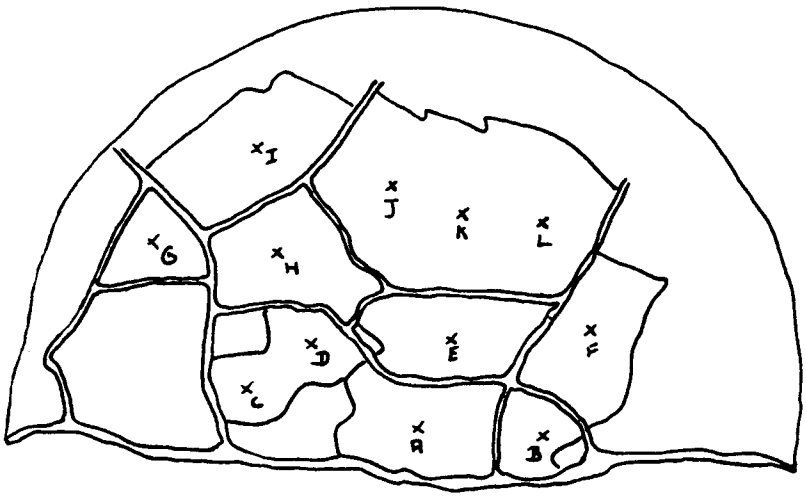
Consequently a very pragmatical path is opted for. The tests set up are miniature problem situations which the subjects are asked to solve either individually or collectively. This simulated problem situation is similar in relevant respects to *actual* problem situations which require for their solution the particular skill that is being tested for. Following this pragmatical route, we developed a particular test for the recruitment of candidate officers, namely, the ROMAT.

The ROMAT is a Group Situational Performance Test. A group of candidate officers, ranging from five to seven persons, are asked to assign sites in a building area on the outskirts of a city for a university, a sporting center, a hotel, etc. Each member in the group is responsible for one such institution. A set of instructions determines what features in a building site are desirable for each institution. There are some instructions all institutions have in common, for example, they all aim at being close to town. One instruction is specific for each institution. In this instruction it is stipulated that the institution in question should be as far as possible from some other institution, for example, the hotel should be as far away as possible from the cultural center, because of the noise problem from loud rock concerts. The subjects in our test are asked to take responsibility for one institution and only they read the instructions only for their own institution. After a few minutes they are asked to take place around a half-circular map of the building area. They are asked to start discussing the location of the institutions for which they are personally responsible as well as of three common projects: a parking lot, an industrial plant, and a storage place for inflammable chemicals. There are eight major and four minor building sites. (In Figure 9.1, B, C, D, and G represent the minor sites.) The institution for which the subjects are personally responsible can only be assigned only to major building sites, the common projects can be assigned to both major and minor sites. After 20 minutes the subjects should come up with some collective agreement.

There is both an objective and a subjective evaluation procedure for the ROMAT. Let us at first consider the subjective evaluation procedure. The test-leader fills out an observation sheet which contains a scale for (1) prominence, (2) efficiency, and (3) sociability, and which also leaves room for more specific remarks on each subject. These scales respectively indicate in how far each subject participates in the group discussion, comes up with good arguments in defense of his or her own interests, and comes up with good arguments in defense of collective interests.

Let us now consider the objective evaluation procedure. The test-leader makes a note of which institutions are ultimately assigned to which sites. Now, all instructions are phrased in terms of optimal distances, for example, the person who is assigned the university knows that this institution should be located as close as possible to town and as far as possible from the industrial plant. The distances between institutions which should be as far as possible from each other are added and subsequently, the distances between

Figure 9.1
The Map of the Building Site in the ROMAT



institutions which should be as close as possible to each other on each person's instructions, are subtracted from this sum. Some distances are measured as the crow flies (for example, the distance to an industrial plant), while other distances are measured via the connecting roads (for example, the distance to town). This procedure yields a score for each person indicating to what respect he or she has been successful in carrying through the instructions. All distances between sites are preprogrammed and scores are computer-calculated. So far no standardization procedure has yet been developed, so only intergroup comparisons between scores are genuinely meaningful.

The Broad-Reductionist Program and the RS1C

We will now turn to the methodological correlate of the broad-reductionist program. The broad reductionist claims that skills can indeed be reduced to more basic components, though this reduction should, aside from personality features and aspects of intelligence, also include modes of cognitive styles. Now the lesson to be learned from the broad-reductionist program for research on the measurement of skills is that our tests should not only focus on the former two variables, but should also have some interest in the latter variable.

We have taken the first steps in developing a computer-steered test which aims at assessing modes of cognitive style, at the same time as assessing aspects of intelligence. This test is essentially a letter substitution test and is named the RS1C. We will at first explain the mechanics of the RS1C and

subsequently report some of the results from our pilot studies. The testtakers are placed in front of a personal computer. The PC explains to them that they will be presented with strings of six letters. Their task consists in reducing these strings to one particular letter by means of eight substitution rules Substitution rules are spelled out in the form XY = Z. A substitution rule, say PO = Q, applies to a particular string, if and only if, this string contains P and O at any place or in any order. Q is then inserted at the location of the first letter substituted for, whether this be P or O. Thus, given PO = Q, the letter string RSOQTP reduces to RSQQT. Subsequently the PC asks our subjects to type in their own names, copy eight strings of three letters and copy the eight substitution rules. Then they are shown how to solve one particular letter string. In the next step they will try to solve one very simple task. The computer teaches them how to make corrections, how to read the record of previously corrected substitutions on the screen, and how to ask for a new task if they decide to give up on the old one. After these instructions the actual test will start. There are six letter strings to be solved. There is no time pressure and the subjects are asked to try to solve all strings.

In a pilot study we checked for the correlations between mean solution time for each subject and results on IQ tests measuring verbal abilities, general reasoning, concentration, memory, and planning. This study yielded negative correlations ranging from $-.23$ to $-.42$. Surprisingly, the poorest correlation was with the concentration test, which was also a letter substitution test, though involving three-letter strings. Apparently letter substitution tasks, involving short versus long strings do not require the same skills for their solution.

Now we also searched for variables in our test which would correlate well with scores in a test measuring modes of cognitive style, namely, Herman's Achievement Motivation Test. Herman's scale expressing debilitating fear of failure correlated significantly with two variables in our own test. There was a significant positive correlation with the copying time of the three-letter strings (.22), of the substitution rules (.37), and with the standard deviation between solution times (.21). There was also a significant negative correlation between copying times for substitution rules and Herman's scale expressing facilitating fear of failure ($-.22$).

For some tasks the solution is more salient than for others in that (1) the order of the letters which should be substituted for in the successive steps on the solution route, is identical in the string and in the substitution rule; (2) these letters occur in the beginning of the string; and (3) these letters are located close to each other. Furthermore, for some task the solution route is the only route one can take on correctly applied solution rules, while for other tasks the solution route is only one alternative among the many branches that all lead to a dead end. We can safely assume that the difficulty of a substitution task is a negative function of the saliency of the correct solution route and a positive function of the complexity of the tree of branches. Now we are trying to construct an index of saliency and an index of complexity for a substitution task. Subsequently, we would like to investigate the following

conjecture. While some persons may be more affected by increasing the complexity of the tree (considering their solution times) for equally salient trees, other persons may be more affected by increasing the nonsaliency (that is, decreasing the saliency) of the task (considering their solution times) for equally complex trees. Now, if this indeed would turn out to be correct, it would be interesting to investigate whether sensitivity to complexity or sensitivity to nonsaliency correlates with more familiar cognitive styles, for example, respectively with debilitating fear of failure and field-dependency. But these are only preliminary ideas for a new research project.

SUMMARY

In the first section, we discussed two situationist objections to personality theory namely that (1) all psychological explanation involving personality traits is merely ad hoc and (2) that there is no constancy to personality traits. Some personality theorists grant that these objections may hold for folkpsychology, but that they can be overcome if we carefully adjust the set of folkpsychological trait names for the purpose of doing proper psychology. We then discussed two such adjustment projects. Pervin (1976,1977) drastically expands the set of folkpsychological trait names by making them situation-specific. Tupes and Christal (1961) (and more recently McCrae and Costa, 1985a, 1985b, 1986, 1987) drastically reduce the set of folkpsychological trait names by isolating five factors in adjective checklists. We argued that the former project supersedes (at least partly) the charge of nonconstancy, though only at the cost of making the charge of ad hoc-ness become more acute. On the other hand, the latter project can (at least potentially) supersede the charge of ad hoc-ness, though only at the cost of making the charge of nonconstancy become more acute. I conclude that the set of folkpsychological trait names (or some close variant) may provide for an optimal balance between both situationist charges and may thus be the best conceptual candidate for psychological modeling. These theoretical ideas are then applied to the development of a free-format self-description method in personality testing.

In the second section we set out three stands in a debate on the ontology of skills: (1) skills are reducible to personality features and aspects of intelligence ("narrow reductionism"); (2) skills are not reducible at all, that is, skills are what they are and nothing less ("antireductionism"); and (3) skills are reducible to personality features, aspects of intelligence, and modes of cognitive style ("broad reductionism"). We went on to show that each ontological stand has its correlate in measurement theory. Then we discussed empirical criterion keying vis-à-vis the assessment of skills in reference to the narrow-reductionist project. The antireductionist stand is illustrated by means of the ROMAT, that is, a group situational performance test for leadership skills which was developed in the Section for

Psychological Research. As a correlate of the broad reductionist stand we presented some preliminary research on the RS1C, a computer-steered letter substitution test, which attempts to measure aspects of both intelligence and modes of cognitive style.

NOTES

1. We have decided not to discuss recent developments in questionaire methods in this chapter. For our own views on this topic, we refer to Claeys et al., 1981. Also worth mentioning is the excellent reader by Angleitner and Wiggins (1986), which covers a broad area in the current research on questionnaire methods.
2. Notice that folkpsychology is not synonymous with pop psychology, that is, the pseudo-psychology that we can find in popular magazines.
3. We have our doubts about the strength of this argument. It certainly is the case that a scoring pattern on the five-personality factors derived from personality features which hold across situations displays more constancy than a scoring pattern derived from highly situation-bound personality features. On the other hand, the problem of constancy still comes in for a person who comes out moderately *agreeable* since she is both highly *suspicious* as well as highly *attentive to people* across situations and consequently describes herself as such.
4. We have added this stipulation to rule out trivial reductions of the form "x is a person who is T (T being some mixed trait)" if and only if "x has the pure cognitive features of being T *and* x has the pure noncognitive features of being T."
5. A reductionist program faces the problem of isolating the components which make up skills. A related project is the study of the complex interrelationships between these components, that is, between personality features, aspects of intelligence and modes of cognitive style. This idea was presented by A. Heim, 1970, pp. 53-61. A. Böhrer and S. Van Den Broucke (1986) have done an extensive study on the correlations between a set of IQ-tests on the one hand and a set of personality tests on the other hand.
6. This conclusion may be slightly too hasty. Consider carefully the scope of the arguments against empirical criterion keying (ECK). The first argument does not only hold against ECK on a narrow reductionistic route, but also against ECK on a broad reductionistic route. Indeed the problem of distinguishing between essential and nonessential features in a scoring pattern also comes in if we bring in variables of cognitive style. The second argument against ECK is actually a more general warning against equating potential skills and actual skills in career-counseling or screening job applicants. Claim (c) against ECK strictly focuses on ECK on a narrow reductionistic route. This claim solely provides for an argument against narrow reductionism if it is indeed the case that ECK for skills on IQ-tests and personality-tests yields relatively poor validity-coefficients. We thus do not intend to take a definite stand in the debate concerning the analysis of skills, since the first two arguments are not strictly arguments against narrow reductionism, while the final claims indeed targets narrow reductionism though is in need of empirical support.

REFERENCES

Allen, B. M., and Potkay, C. R. (1973). Variability of Self-Description on a Day-to-Day Basis: Longitudinal Use of the Adjective Generation Technique. *Journal of Personality, 41*, 638-652.

———. (1977). The relationship between AGT Self-Description and Significant Life Events: A Longitudinal Study. *Journal of Personality, 45,* 207–219.

Anastasi, A. (1961). *Psychological Testing* (2nd ed.). New York: Macmillan.

Angleitner, A., and Wiggins, J.S. (eds.). (1986). *Personality Assessment via Questionnaires—Current Issues in Theory and Measurement.* Berlin: Springer Verlag.

Böhrer, A. (1980). Social Anxiety, Self-Confidence and Military Leadership. *Proceedings of the 22nd Annual Conference of the Military Testing Association, 1,* 91–102.

Böhrer, A., and Van Den Broucke, S. (1986). Personality and Test Performance. In S.E. Newstead, S.H. Irvine, and P.L. Dann (eds.), *Human Assessment: Cognition and Motivation.* Dordrecht, Martinus Nijhoff, 287–298.

Claeys, W., De Boeck, P., and Böhrer, A. (1981), How to Improve the Validity of Personality Questionnaires. *Proceedings of the 23rd Annual Conference of the Military Testing Association, vol. 1.* Coordinated by the U.S. Army Research Institute for the Behavioral and Social Sciences, Arlington, Virginia.

Claeys, W., De Boeck, P., Van Den Bosch, W., Biesmans, R., and Böhrer, A. (1985). A Comparison of One Free-Format and Two Fixed-Format Self-Report Personality Assessment Methods. *Journal of Personality and Social Psychology, 49*(4), 128–139.

Elshout, J. J., and Akkerman, A. E. (1975). *Vijf-Persoonlijkheidsfactoren Test (SPFT) Handleiding.* Nijmegen: Berkhout.

Heim, A. (1970). *Intelligence and Personality—Their Assessment and Relationship.* Harmondsworth: Penguin.

Hermans, H. J. H. (1976). *P.M.T. Prestatiemotivatietest—Handleiding.* Amsterdam: Swets & Zeitlinger.

Le/De Romat (1986). Manuel pour les observateurs/Handleiding voor Observatoren. Section for Psychological Research (Belgian Army), mimeo.

McCrae, R. R., and Costa, P. T. Jr. (1985a). Comparison of the EPI and Psychoticism Scales with Measures of the Five-Factor Model of Personality. *Personality and Individual Differences, 6,* 587–597.

———. (1985b). Updating Norman's "Adequate Taxonomy": Intelligence and Personality Dimensions in Natural Language and Questionnaires. *Journal of Personality and Social Psychology, 49*(3), 710–721.

———. (1987). Validation of the Five Factor Model of Personality Across Instruments and Observers. *Journal of Personality and Social Psychology, 52*(1), 81–90.

McCrae, R.R., Costa, P.T. Jr., and Busch, C.M. (1986). Evaluating Comprehensiveness in Personality Systems: The California Q-Set and the Five Factor Model. *Journal of Personality, 54*(2), 430–446.

Molière, J.B. (1947 [1673]). *Le malade imaginaire.* Bruxelles: Les Editions de la Cité Genève.

Pervin, L.A. (1976). A Free-Response Description Approach to the Analysis of Person-Situation Interaction. *Journal of Personality and Social Psychology, 34*(3), 465–474.

———. (1977). The Representative Design of Person-Situation Research. In D. Magnusson and N. Endler (Eds.), *Personality at the Crossroads: Current Issues in Interactional Psychology.* New York: Wiley, 371–384.

Potkay, C. R., and Allen, B. P. (1973). The Adjective Generation Technique: An Alternative to Check Lists. *Psychological Reports, 32,* 457–458.

Tupes, E. C., and Christal, R. E. (1961). *Recurrent Personality Factors Based on Trait Ratings.* USAF ASD Technical Reports (No. 61–97).

Vrije Zelfbeschrijvingsmethode. (1987). Lisse (Netherlands): Swets Test-Services. In translation, *Free-Format Self-Description Method.* (1987). Kortenberg (Belgium): Center for Basic Interactive Research. *L'auto-description libre.* (1987). Kortenberg (Belgium): Center for Basic Interactive Research.

Willems, L. F. M., Tuender-de Haan, H. A. and Defares, P. B. (1973). Een Schaal om Sociale Angst te Meten. *Nederlands Tijdschrift voor Psychologie, 28,* 415–422.