Special Report PERS-423 1987

Computer technology and the prospects for innovation in personality assessment

Martin J. Terre Blanche

Nasionale Instituut vir Personeelnavorsing — NIPN
National Institute for Personnel Research — NIPR

RGN·HSRC

Computer technology and the prospects for innovation in personality assessment

Special Report PERS-423

Computer technology and the prospects for innovation in personality assessment

M.J. Terre Blanche

Pretoria
Human Sciences Research Council
1987

Martin J. Terre Blanche, B.A. Hons., Researcher

National Institute for Personnel Research
Executive Director: Dr G.K. Nelson

ISBN 0 7969 0540 1

Printed and published by HSRC
134 Pretoriusstreet
Pretoria

ACKNOWLEDGEMENTS

Dr G.K. Nelson : Executive Director, NIPR

Dr T.R Taylor : Head, Test Construction Division

Charles Chemel : Psychometric Methods Division

NIPR library staff

CONTENTS

FIGURES

ABSTRACT

In this report the potential impact of computer technology on personality assessment is discussed. Technology can be said to play a mediational role between tester and testee. There are three dimensions to this role and personality tests can be classified in terms of these. The first concerns the degree of interactivity permitted by the technology; the second refers to the way in which the tester "delegates" his or her assessment function; and the third is the degree of subtlety of the test material. Computer technology has its clearest impact on the first dimension, namely interactivity. To make interactive assessment possible changes would have to occur in the psychometrics and personality theory commonly associated with personality assessment. An interactive personality test is presented as an illustration of the possibilities and difficulties involved in computerized personality assessment.

EKSERP

In hierdie verslag word die invloed van rekenaartegnologie op persoonlikheidsmeting bespreek. Tegnologie funksioneer as tussenganger tussen toetser en toetsling. Die tussenganger-funksie bestaan uit drie dimensies en persoonlikheidstoetse kan hiervolgens ingedeel word. Die eerste dimensie het te make met die graad van interaktiwiteit wat deur die tegnologie moontlik gemaak word; die tweede verwys na die manier waarop die toetser sy of haar toetsingsfunksie "delegeer"; en die derde is die mate waartoe toetsmateriaal subtiel voorkom. Rekenaartegnologie beinvloed die eerste dimensie, naamlik die graad van interaktiwiteit, die meeste. Interaktiewe toetsing is slegs moontlik indien veranderings plaasvind in die psigometrika en persoonlikheidsteorie wat met persoonlikheidsmeting verband hou. 'n Interaktiewe persoonlikheidstoets word beskryf ter illustrasie van die moontlikhede en probleme wat met gerekenariseerde persoonlikheidsmeting saamgaan.

1. INTRODUCTION

Personality assessment, traditionally the poor relation of the psychological testing enterprise, can be seen to have reached a point of stagnation. The "golden age" of personality assessment during the 1920s, 30s and 40s, when objective questionnaires, the inkblot technique and thematic apperception were invented, has given way to a period informed by a spirit of refinement rather than innovation.

In this regard Lanyon and Goodstein (1982) state: "Further improvements in personality assessment will probably have to come from new approaches to the problem rather than from further refinements of the traditional methods" (p. 42). Three potential sources of innovation are: Personality theory, psychometrics and new technology.

Personality theory is perhaps the most obvious potential source of innovation in personality assessment. During the last forty years there has been a gradual shift in personality theorizing away from the intrapersonal towards the interpersonal, so that there is now a marked discrepancy between the view of personality implied by most personality measures and currently fashionable theory. Although this discrepancy no doubt does exert some pressure towards innovation, one should remember that the mapping from theory to assessment methodology is fairly indirect. The self-report questionnaire methodology, for instance, is associated primarily with a trait view of personality, but has nevertheless been used in theoretical milieus as diverse as

traditional psychiatric nosology (Hathaway & McKinley, 1951), psycho-analytic theory (Grygier, 1956), and interpersonal theory (Schutz, 1977). Thus it would seem that personality theory in itself is unlikely to provide sufficient impetus for radical reform in personality assessment.

The second potential source of innovation, psychometrics, also seems unlikely to provide the necessary impetus. The strict psychometric approach is in fact allied to the tradition of refinement rather than innovation. Much time and energy is spent on rigorous statistical analysis, as can be seen in the voluminous literature on response styles and sets (reviewed in Kleinmuntz, 1967) or the ever-more refined methodologies of item selection (Burisch, 1986). Wider implications for the structure of personality tests as such are seldom drawn. Some of the historical reasons for this are discussed in Chapter 4.

The third potential source of innovation, and the one considered in this report to be the most likely catalyst for change, is technology. Technology represents as much of an input to the test construction process as psychological theory or psychometrics. One could argue, for instance, that the present age of mass testing would not have been possible but for the invention of the printing press and the concomitant exponential rise in the literacy rate. On a more mundane level, new assessment techniques have sprung from technologies created for other purposes. The Rorschach technique, for instance, began

life as an experiment in visual perception (Schachtel, 1966), and self-report questionnaires evolved from attempts to construct standardized interviews (Lanyon & Goodstein, 1982).

The technology of the moment is of course the digital computer. Dire warnings (Dreyfus & Dreyfus, 1986) and utopian predictions (Simon, 1984) as to how computers are about to affect our lives abound, and whichever of these prognostications proves to be correct, it seems virtually certain that many more people are due to feel the impact of computers in the near future. In psychology also, and in the field of personality assessment, computers seem set to become an important factor. One indication of this is that most authors of introductory texts on assessment now include sections on computerization in their works (see Cronbach, 1984; Lanyon & Goodstein, 1982). An overview of the state of the art in computerized personality testing is given in Chapter 2.

There is a prevailing mood of optimism about the effect of computers on assessment. The uses to which computers can be put are thought to be virtually limitless, and there are speculations about a brave new world of automated but individualized assessment (see for instance Erdman, Greist, Klein, Jefferson & Getto, 1981). In this report a more cautious approach is taken. It is argued that technology plays a mediational role in assessment and that the potential impact of computer technology can best be analysed in these terms. The approach is developed in Chapters 3 and 4. In Chapter 5 a new personality test that makes fuller use of computer technology is described.

2. THE STATE OF THE ART

In this chapter a short overview is given of previous work in the field of assessment by means of computers, the intention being to indicate how the present study fits into its historical context. Specific attention is paid to computerized personality assessment.

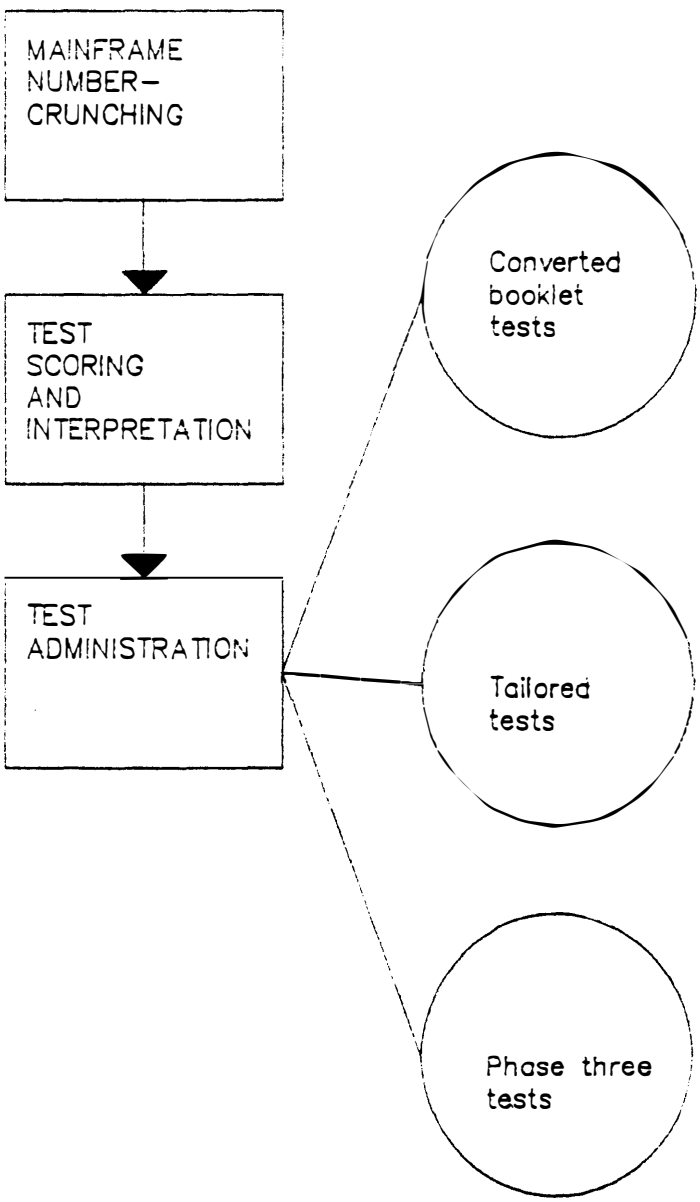
2.1 Computers in assessment: An historical perspective

A schematic representation, based on Lanyon and Goodstein (1982), of the history of computerized personality assessment is given in Figure 2.1. As can be seen, the penetration of computers into psychological assessment can be represented as having occurred in three ways. First, large mainframe computers began to be used to perform the data analysis required in the construction of assessment instruments. The factor-analytic approach, which has been such a feature of the assessment landscape, would probably never have become as popular as it has if computers were not available.

Secondly computerized test scoring and interpretation services were developed. Although South African companies have not taken to computer scoring and interpretation in a large way, these services have become multimillion dollar business in the United States. This use of computer technology is more controversial than the first-mentioned, because the commercial enterprises who

market these services are often interested in presenting a slick-looking product rather than in scientific integrity (Lanyon, 1984).

Figure 2.1: An historical perspective on computers in assessment



The third way in which computers have come into the assessment landscape is still in its infancy. This is the use of computers to administer tests. The growth of computerized test administration is occurring in three phases: firstly pencil and paper tests are converted for computer administration, secondly

so-called "tailored" or "adaptive" tests are constructed, and finally entirely new kinds of tests that make full use of the computer's capabilities are constructed.

A very large number of pencil and paper tests have now been converted for computer administration. In South Africa the NIPR has converted the majority of its own tests as well as some of those developed by IPER for administration on the PLATO psychological testing system (Taylor, Gerber & Rendall, 1981). Although the advantages of this kind of testing are many, these do not begin to exploit the capabilities of the computer.

Adaptive or tailored tests make fuller use of the computer's capabilities. These tests (e.g. those of Niedner & Fink, 1981; and Watts, Baddeley & Williams, 1982) are so designed that the computer presents a uniquely selected subset of items to each testee, and is able from this to compute scores equivalent to those obtained from a full-length test. The technique used is that of repeated "branching". The most appropriate item to present next is selected on the basis of information gleaned from previous responses. Adaptive testing is often based on very sophisticated statistical models (reviewed in Vale, 1981), but typically differ from conventional tests only in the speed with which they can be administered.

The third phase of computerized test administration involves the construction of novel kinds of tests that make full use of computer technology. Phase three computerized personality testing is the focus of this report.

2.2 Previous research on computerized personality testing

In the field of cognitive testing, adaptive procedures are now fairly commonplace and some incisive analyses of further opportunities for innovation have been published (e.g. Hunt & Pellegrino, 1984; Taylor, 1987). Attempts at constructing adaptive personality tests are still comparatively rare, however. A few more-or-less qualifying examples are described by Johnson et al. (1981), Kleinmuntz and McLean (1968), and Lyons and Brown (1981). Analyses of the effects of computer-administration on the future of personality assessment (e.g. Space, 1981) also tend to be less wide-ranging than equivalent efforts in the cognitive field. Many authors are merely dazzled by the seemingly limitless magic of computers. Even Cronbach (1984) appears to fall into this category when he blithely claims that computers "can simulate a human interviewer remarkably well" (p. 40). Erdman et al. (1981) display signs of the same kind of computer euphoria when they predict that computers will take over from psychiatrists in the near future.

Two aspects in particular of the computerized personality assessment situation seem to have captured the imagination of writers on the topic: The flexibility of computerized

questionnaire administration and the impact that the computer as a social stimulus might have on a testee's responses.

2.2.1 Flexibility of administration

Many authors use an analogy with interviewing to describe the potential flexibility of the computerized administration of questionnaires. Stout's (1981) analysis is perhaps the most sophisticated. He describes a human interviewer as basically following a fixed schedule, but with the flexibility to interrupt if necessary (e.g., to explain a question, ask for more co-operation, or ask the interviewee if he or she is tired or distressed). There are, in effect, two interviews: a lowerlevel interview and a meta-interview. Stout suggests that a computer questionnaire could be programmed to have a meta-interview function, capable of interrupting the lowerlevel questionnaire in order to insert extra items, tell the testee to rest, ask for cooperation or call a human administrator.

Erdman, Klein and Greist (1983) see the advantage of flexible computer administration as follows: "Branching logic that is not possible with a paper-and-pencil questionnaire enables the interview to flow naturally, concentrating on relevant questions, picking up missed information, and even tailoring future questions to the respondent's particular characteristics" (p. 66). Johnson et al.'s (1981) Psychological Systems Questionnaire, a sprawling computerized psychiatric intake

interview, is a concrete example of the kind of instrument that can be constructed when an attempt is made to simulate the flexibility of human interviewing.

Technically, the advantage of flexible questionnaire administration is that it combines "bandwidth" with "fidelity", unlike conventional tests where the one usually has to be traded off against the other (Weiss, 1982). A conventional omnibus personality questionnaire has considerable bandwidth in that it has items covering a wide area, but low fidelity in that only a few items are truly appropriate for any particular testee. 'Tailored' tests, on the other hand, are designed to present each testee with a short subset of appropriate items from a large universe of items, thus combining bandwidth and fidelity. The ability to maintain bandwidth without sacrificing fidelity represents an improvement in degree, not in kind. The intention appears still to be to get the same kinds of results as with conventional questionnaires, only more efficiently. No systematic analysis, as is attempted in this report, as to how the flexibility of computer administration could be used to obtain entirely different kinds of personality information could be found in the literature.

2.2.2 The social impact of computers

The social impact computers may have on the way testees respond to personality questionnaire items is the second topic that has received some research interest. Space (1981) puts it succinctly

when he points out that "like a human tester, the computer has a set of demand characteristics and is not necessarily neutral in the testing situation" (p. 603). Many other authors share similar sentiments (e.g. Bartram & Bayliss, 1984; Beaumont, 1981; Johnson & Johnson, 1981), but unfortunately only two aspects of the computer's alleged non-neutrality has received systematic attention: The degree of honesty computers elicit from testees, and the degree of rapport they establish with testees.

Koson, Kitchen, Kochen, and Stodolosky (1970), in an early study of the degree of honesty elicited by computers, hypothesized that testees would be most honest when answering booklet-administered MMPI items, second most honest when answering computer-administered items, and least honest when the items are presented by a human tester. They based this hypothesis on the belief that the greater degree of anonymity offered by computers and booklets as compared to humans would encourage testees to be more honest. Koson et al.'s results were in the expected direction, but not statistically significant. Since then several similar studies (reviewed by Space, 1981) have come up with significant results, although there have also been some results pointing in the opposite direction (Erdman, Klein & Greist, 1983).

The anonymity-equals-honesty theory cannot be taken too seriously. It is, in fact, possible to argue for the exact opposite, as do Liebert and Spiegler (1982): "Self-report inventories...are particularly susceptible to dissimulation inasmuch as the absence of an examiner lends a somewhat more

impersonal and distant character to the evaluation" (p. 284). In addition, the degree of anonymity associated with a computer is not a fixed property, but depends, among other things, on evolving cultural attitudes towards computers. With the growing availability of computers - Hasset in 1984 estimated that 86 % of American high schools had one or more computers - it is possible that computers as such will attract more neutral attitudes. The issue may become not whether to trust the computer with one's innermost secrets, but whether to trust the agency behind the computer (Thompson, 1984). In the next chapter an attempt is made to develop a classification of the way in which the person behind the computer manifests himself or herself, through the computer, to the testee.

The literature dealing with the degree of rapport that can be established between a computerized test and a testee is more sophisticated. Acknowledgment is given to the fact that the computer's social impact is not a fixed property, but depends on the manner in which the test is presented. Although studies of the sort that set out to prove that merely using a computer to administer any test already makes for better rapport with testees can also be found (e.g. Erdman et al., 1983), there is more generally a realization that rapport "is very much dependent on the design of the program software" (Bartram & Bayliss, 1984).

Bartram and Bayliss (1984) suggest that a high level of rapport will be attained if the machine's behaviour is consistent with the user's internal model of it. In the same vein as Bartram and Bayliss, Beaumont (1981) speculates that it would be

"disconcerting for a testee... to feel uncertain of the outcome of actions in terms of the machine's response" (p. 431). Johnson and Johnson (1981) cite evidence that "users feel worse about interactive computer systems when they lack control over them" (p. 423). In the next chapter the degree of control a testee has over the consequences of his responses is discussed in terms of a dichotomy analogous to that found between dreams on the one hand and games on the other.

Apart from the predictability of the programme, another factor that has been thought to affect rapport is the degree of intimacy with which the programme addresses the testee. Beaumont (1981) believes that there is "a difficult balance to be struck between a system that is cold and impersonal and one that is falsely human" (p. 431). Computerized counselling programmes vary as to the degree of intimacy they assume towards the testee. Selmi, Klein, Greist, Johnson and Harris (1982) use unequivocally anthropomorphic cues, even down to the computer introducing itself by name. Wagman (1982), on the other hand, opts for a strictly impersonal tone. In the next chapter the degree of intimacy is discussed in terms of a dichotomy between programmes presenting themselves as "beings" <u>vs.</u> programmes presenting themselves as "worlds".

A final issue relating to the degree of rapport possible between a human and a computer concerns the fact that a computer is for all intents and purposes disembodied. Much of the considerable literature on human-human rapport (reviewed by Cook, 1983) concerns interpersonal messages passing between interactants on

the non-verbal level, and the assertion is often made that non-verbal messages are of primary importance in establishing rapport. This of course represents something of a problem for a computerized test, trapped as it is in a verbal world. There are two ways of getting past this problem. The first is to say that computers are in fact capable of receiving and transmitting non-verbal messages. Stout (1981) describes how a computer system could be relatively easily set up to receive non-verbal data such as heart rate, eye movement, force of hitting the keys, response latencies and even body posture. On the transmitting side, Johnson & Johnson (1981) cite evidence that essentially non-verbal messages from the computer, such as the absolute time taken to respond and the variance in time, strongly influence the user's attitude. As an extreme measure an all-purpose interviewing robot using voice recognition (Richards, Fine, Wilson & Rogers, 1983) and a programme for interpreting responses to open-ended questions (Space, 1981) could perhaps at some future date be constructed.

I shall adopt the position that non-verbal communication is not all that important in establishing rapport. Support for this position can be found in Krauss, Apple, Morency, Wenzel and Winton (1981) who maintain that the importance of the non-verbal channel has been exaggerated, even in emotional communication, and that the "display rules" governing a particular situation determine which channel is most important. Since people do not expect to interact with a computer in non-verbal terms - i.e. it

is not written into the display rules - they will send and receive messages primarily via the verbal channel.

2.3 Conclusion

It has been argued in this chapter that the penetration of computers into assessment has occurred in three ways, and that the third way (computerized test administration) has itself occurred in three phases - the third phase being the construction of innovative computerized tests. Previous research on phase three personality assessment has been found to be limited, consisting mostly of speculations on how the flexibility of computerized administration and the social impact of the computer might beneficially affect the quality of assessment.

3. ASSESSMENT AS A RELATIONSHIP MEDIATED BY TECHNOLOGY

In the previous chapter it was stated that past work on computers as a force for innovation in personality assessment is limited. An attempt is made here to develop an alternate theoretical base. The potential impact of computers on personality assessment is evaluated in terms of the mediational role of assessment technology and a model of assessment as a mediated relationship is presented. The literature drawn on in this chapter includes not only that dealing with personality assessment, but also human-computer interaction and other fields. As a consequence the tone and content of the chapter is somewhat different from that usually associated with work on personality assessment. It should be borne in mind that in this chapter assessment is viewed from an unusual angle, namely that of the technology involved. In Chapter 4 the focus is again narrowed down to fields more commonly associated with personality assessment, namely psychometrics and personality theory.

3.1 Assessment as a relationship

It is not difficult to find support for the notion that personality assessment involves, at the most basic level, a relationship between assessor and assessee. Cronbach (1984), Hunt (1980), Lanyon and Goodstein (1982) and Sundberg (1977), to name but a few, advocate viewing assessment as a form of social

relationship. However, in practice testees have most often been treated not as partners in a relationship, but in isolation as "stimulus-response mechanisms" (Hess & Neville, 1977, p. 170). An analogous situation exists in experimental psychology where, ever since Silverman's (1977) seminal work, the call has consistently been for viewing experimentation as a social relationship, although in practice the asocial natural science model has prevailed (see the criticisms of Davis & Perkovitz, 1979; Duncan, Kanki, Mokos & Fiske, 1984; Kenny & LaVoie, 1985; Warner, Kenny & Stoto, 1979; Wright, Blackmer & Ingraham , 1984; and Wright & Ingraham, 1985).

The "mechanistic" and the "relationship" views of what goes on in assessment have been given various different labels. Tetlock and Manstead (1985) use the terms "intrapsychic" and "impression management" to identify the two opposing ways of viewing a testee or experimental subject. Johnson (1981) uses the terms "self disclosure" and "self presentation". The self disclosure (or intrapsychic) view assumes that test responses are simple factual communications or signs that refer unambiguously to the self. Responses are, in this view, valid because they accurately report or correspond to underlying traits that shape future behaviour. The self presentation (or impression management) view, on the other hand, assumes that responses are "a picture of how one generally would like to be regarded" (Johnson, 1981, p. 763). In this view responses are valid because the style of self-presentation which a person employs in a test or experiment is similar to his or her style in everyday interaction.

Some rather convincing research has been conducted (Johnson, 1981) demonstrating the superior predictive power of viewing responses to a personality questionnaire as a form of self-presentation or impression management. The tolerance with which social desirability responding (a form of impression management) is these days regarded (e.g. by Angleitner & Wiggens, 1986; and Jackson and Helmes, 1979) further attests to the growing acceptance of the impression management view. To the author's knowledge, however, no systematic model of personality assessment as a social relationship has as yet been published. It is not unlikely that the lack of such a model is at least in part responsible for the fact that assessment has in practice remained stuck in the intrapsychic paradigm.

3.2 The human-computer relationship

There is frequent reference in the literature on computer use to the idea that human-computer interaction can be viewed in relational terms. At the least, as was discussed in Chapter 2, computers can be said to have a particular social impact; they may draw certain attitudes from people such as perhaps an honesty set or a feeling of mutual rapport. As was discussed in Chapter 2, the disembodied nature of computers presents a problem when concepts from human interaction, such as non-verbal communication are applied to human-computer interaction, but these problems are not necessarily insurmountable.

In Artificial Intelligence (AI) research the use of an analogy with human conversation to describe human-computer interaction has become commonplace (Allen, 1982). Sackman (1967) lists in chronological order developments in computer technology which have affected human computer communication: "Hardwiring" (the computer has to be rewired to make it perform differently); communication by means of punched cards and esoteric programming languages; and communication by means of easy-to-use screen displays, keyboards, lightpens etc. This historical progression from slow, technically demanding communication to direct, naturalistic communication in itself increasingly suggests an analogy with human interaction.

In discussing the degree to which human-computer interaction resembles human interaction, Pinsky (1983) comes to a conclusion which anyone who has interacted with a computer is likely to sense as correct, namely that human-computer interaction "has an ineradicably double nature: it is at once both conversational and mechanical" (p. 38). While intuitively agreeing with this conclusion, one might wonder why the dual nature of the interaction need be "ineradicable". Surely advances in computer capabilities could rid the interaction of its mechanical quality? This puts one in the midst of the controversy on whether machine intelligence is possible. If it is, as the pro-AI lobby (e.g. Feigenbaum & McCorduck, 1984) maintains, then the present 'mechanicalness' of human-computer interaction is merely an engineering problem. If it is not, as the antilobby maintains (e.g. Dreyfus & Dreyfus, 1986), then the problem is indeed ineradicable.

An alternative approach, which sidesteps the question of machine intelligence, is proposed by Maass (1983). Maass maintains that mechanical, stilted interactions are most often the consequence of delegation. This is analogous to the interaction that one may have with an inflexible and ultimately unhelpful clerk at an information desk, who has had his particular role delegated to him by some higher authority. According to Maass the role of interaction partner has been delegated to the computer by (for instance) the programmer; the computer is then a "virtual communication partner" whose behaviour will inevitably have an element of mechanicalness about it.

Maass (1983) sees the computer as an intermediary in what is essentially a relation between humans. In other words he goes beyond an analogy with human-human interaction to suggest that <u>real</u> interaction between humans is implicated when a human interacts with a computer. Maass' approach is followed in this report and appears to be finding increasing support elsewhere as well. Bobrow and Hayes (1985), for instance, advocate that the human origin of and responsibility for the behaviour of AI programmes be made explicit, while Coulter (1985) argues that computer programmes only refer to the real world in a way that is parasitic on the programming language, which in turn is parasitic on the programmer behind the computer. Seen from the other side, programmes can also be said to refer to the real world in a way that is parasitic on the <u>user.</u> Good (1985) maintains that AI programs may appear intelligent only by virtue of what is "read into" their behaviour by users.

To summarize: There is strong support for viewing personality assessment in terms of the assessor-assessee relationship, but this has not been put into practice - possibly because the terms in which it could be done has not been made explicit. Human-computer interaction can also be viewed as a form of social interaction, although it is characterized by a certain stilted quality. A particularly interesting way of accounting for this is to view the computer as an intermediary in a relationship between humans.

3.3 A model of assessment as a mediated relationship

Having established that assessment can be viewed as a relationship and that computers can be viewed as mediating human relationships, the scene is set for the development of a model of (computerized) assessment as a mediated relationship. The point of departure for the model is to view personality assessment in terms of the relationship between assessor and assessee. A peculiar characteristic of assessment relationships is that they are transacted via some form of testing technology. Thus the assessment relationship referred to here is not the incidental social pleasantries that inevitably pass between test administrator and testee, but the relational interplay between some real or imagined tester and the testee <u>by means of the test.</u> The model does not seek to incorporate particular extratest relational contingencies such as a desire by the one party to fake and by the other to prevent faking, but only such

relational circumstances as can be deduced from the structure of the test itself. The test is seen as a mediator and testtaking is seen as a mediated experience.

The model is no doubt not complete and perhaps somewhat arbitrary and it is to be hoped that in time more sophisticated models will be developed. The one virtue claimed for it is that it is more structured and specific than the rather general statements about the desirability of viewing personality assessment in relationship terms that have thus far characterized the literature.

The model consists of three orthogonal dimensions, labelled interactivity, delegation, and intentions. The dimensions may also be described (respectively) as follows: Dreams/games, beings/worlds, and questions/koans. The second set of rather informal labels were chosen in consideration of the fact that the model is not intended as a final theory, but as a working model which should be flexible and open to amendment.

3.3.1 Interactivity (dreams vs games)

In a face-to-face assessment situation, the assessor, being human, cannot but react to the assessee; the assessee therefore feels immediately the consequences of his or her responses. The adept assessee in such a situation, say an employment interview, knows how to juggle half a dozen considerations to arrive at the most pleasing next response: How best to capitalize on the

interviewer's reactions in order to "fine tune" responses; how much credence to attach to a particular reaction from the interviewer; how far to stretch the truth to make it sound like something the interviewer would like to hear; etc. The interviewer, meanwhile, may be playing a game of his own - that of second-guessing the assessee's game, and so on.

In a mediated situation, such as the test-taking situation, the position is different. Feedback need not be immediate, but may be deferred; the testee may only be given feedback after the completion of the test, or (as is unfortunately more often than not the case in personality testing) feedback may be deferred indefinitely. Thus face-to-face assessment is inevitably interactive or game-like, but mediated assessment may be "reactive" or dream-like. A reactive (or conventional) test can be likened to a dream in the sense that all interactions that occur are, inevitably, the dreamer's own inventions. A dreamer is in full, if not conscious, control of the script of his or her dream. In a dream-like test the assessor's immediate reactions are not available to the testee and can at most be imagined. This kind of "stonewalling" has its uses, of course, but must also result in testees assimilating the tone of the interaction and responding in a lifeless and impoverished manner.
The relative lack of predictive validity of conventional personality questionnaires (Mischel, 1968) may in no small part be due to the dream-like quality of subjects' responses.

With the advent of computerized testing this is set to change. Computers make interactive assessment, analogous to the face-to-face assessment described above, possible. Whatever the other implications are of a move from booklet to computer technology, the fact that interactive assessment now becomes a possibility is clearly crucial. It is for this reason that interactivity is included as the first dimension of the model. An entire chapter will moreover be devoted to the impact interactive assessment is likely to have on psychological and psychometric theory.

A good starting point for coming to a clearer understanding of what interactive assessment involves is the idea of games. This will be explored below with special reference to the literature on human-computer interaction. In the next chapter more traditional sources (personality theory and psychometrics) will be drawn on to explicate the concept of interactivity further.

It can be argued that personality assessment is in some ways a game-like activity and is likely to become more so as computerized assessment gains ground - but what sort of game is it that is played between assessor and assessee? Kelly's (1958) witty but insightful remark concerning the difference between objective and projective personality tests begins to provide the answer: "When the subject is asked to guess what the examiner is thinking, we call it an objective test; when the examiner tries to guess what the subject is thinking, we call it a projective device" (p. 332). The assessment game, it is suggested, is the familiar interpersonal game of "guess what I'm thinking".

"Guess what I'm thinking" operates differently in dream-like and in game-like tests. In dream-like tests, such as conventional personality questionnaires, the testee has to guess what the tester thought (and vice versa). To do well in this kind of test one needs to be "testwise". Hess and Neville's (1977) definition of testwiseness suggests why this is so: "A testwise individual is able to modify test responses and results to conform to his/her hypotheses about the test" (p. 170). Put differently, a testwise person is able to capitalize on the subtle cues contained in the test material through which the assessor's theories about people are revealed. The testee's "success" depends on swiftly evolving a model of the assessor's model of what people are like and then adapting his or her responses accordingly.

In an interactive personality test testwiseness may not be sufficient for "successful" performance. To do well, a testee must be able to capitalize on the subtle cues relevant to the current "hypotheses" about him or her held by the test. Success depends, therefore, on evolving a model which "tracks" the test's evolving model of him or her, which in turn is based on the assessor's static model of what people are like.

The analogy to games to describe interactivity has been used before in the field of human-computer interaction, for instance by Pinsky (1983) who refers to what transpires between a human and a computer as a "language game" of some sort. An exact parallel to the two kinds of models discussed above exists. A

much-used concept in Artificial Intelligence (AI) research is that of "user modeling" (Gilbert & Heath, 1985). Most computer systems that interact with humans "maintain, albeit only implicitly, a model of their users" (Boguraev, 1985, p. 127). In other words, the system expects its users to have certain requirements and to act in certain predictable ways. Boguraev calls user models of this kind, i.e. models that represent "an abstraction of a general class of rules" (p. 127), empirical user models since they are often empirically tested. This is the equivalent of the assessor's "static model" of people discussed above.

The kinds of user models that AI workers hope to incorporate in their programmes go beyond empirical models, however. These kinds of models consist of inferences drawn by the programme about its current user's goals, plans and concerns (Good, 1985) and are known as dynamic user models (McTear, 1985). In psychometric terms this is analogous to the test's 'evolving model' of the testee, referred to above.

Two interesting properties of dynamic user models are firstly that they can only be generated with reference to the context and secondly that it is virtually impossible not to get stuck in an infinite regress when discussing them. The first point, that context is the source of a dynamic user model, is made as follows by Boguraev (1985): "As there are typically no visual or other sensory data available to the system, it must rely entirely on the sequence of typed user inputs up to the current point in the

exchange, interpreted in the context of the running dialogue" (p. 127). In Chapter 5 it will be shown how context can be used in an actual test to form and maintain a "testee model".

The other point, concerning the infinite regress, is also obvious, once stated. If a system's model of a user is to be complete, it must also model the user's model of the system, which, in turn, must be assumed to model not only the system, but also the system's model of the user's model of the system, and so on ad infinitum.

The infinite regress problem is not unfamiliar in AI work. Much of AI's still unfulfilled promise in areas such as natural language translation and robotics hinges on the difficulty of interpreting facts in context, and any appeal to context automatically invokes the infinite regress problem. Dreyfus (1979), one of the fiercest critics of AI, makes the problem explicit in this antimony: "On the one hand, we have the thesis: there must always be a broader context; other-wise, we have no way to distinguish relevant from irrelevant facts. On the other, we have the antithesis: there must be an ultimate context, which requires no interpretation; otherwise, there will be an infinite regress of contexts, and we can never begin our formalization" (p. 222). Compounding the infinite regress problem is a vicious-circle problem that goes with it: Any holistic, contextual interpretation of reality is bound to get caught up in a vicious circle, where the elements get their meaning from the context and the context can be defined only in terms of its elements (Bateman, 1985).

Dreyfus (1979) suggests that humans, unlike computers, can transcend the regress and the vicious circle because they are always already-in-a-context and because they can appeal to an ultimate context, which he calls the human-life-world, from which all else can be deduced.

Another way out is suggested by the explicitly holistic tradition embodied in systems theory. While the theory concedes that every lowerlevel system is an element, or subsystem, in a higherlevel system (Jordaan & Jordaan, 1980), a regress and vicious circle is avoided by the simple fiat of stating that the choice of level to focus on is ultimately arbitrary - there are always bigger or smaller systems one might have focused on. No one system typifies the truth: One's choice of system to focus on is governed by the sorts of information one hopes to obtain. A term sometimes used by systems psychologists to express a similar concept is "punctuation", which is an arbitrary and "artificial break into sequences of behaviour in a variety of contexts" (Campbell, Reder, Draper & Pollard, 1982, p. 45).

Looked at from this perspective, the very short behaviour sequences focused on in traditional psychometrics is a perfectly legitimate form of punctuation, although not necessarily the one that will yield the most useful information on 'personality', nor necessarily the one best suited to exploiting the capabilities of computers. In Chapter 4 the psychometrics of a different form of punctuation will be explored.

In summary, it has been suggested that the first way in which the mediational role of technology may be viewed is in terms of interactivity. Computers make interactive assessment possible, which suggests that computerized tests could become more game-like than conventional tests. The fundamental game that is played in an interactive personality test is that of guess-what-I'm-thinking - a game that involves the construction of what may, by analogy to AI research, be called dynamic user models by both the test and the testee. This kind of reciprocal modeling invokes an infinite regress, which is probably best circumvented by arbitrary "punctuation". The punctuation recommended in this report is at the level of assessor-assessee or test-testee interaction, but the psychometric and personality-theoretical implications of this need to be investigated further.

3.3.2 Delegation (beings vs worlds)

The concept of delegation is borrowed from Maass (1983) who, as discussed earlier, views programmers as delegating their communicating function to programmes, which consequently operate as virtual communication partners. In the present context the equivalent is to say that a personality test interacts as a virtual communication partner with a testee, this function having been delegated to the test by the assessor.

Thus the second way in which a personality test may mediate the interaction between assessor and assessee is through its function as a virtual communication partner. There are two ways in which this function may be performed, i.e. there are two forms of delegation. The first is where the test presents as a single human-like entity, an artificial being, functioning as a more or less direct representative of the assessor. The second is where the test presents as an entire situation, an artificial world, perhaps populated by numerous human-like entities. The first case is analogous to an executive delegating to a clerk, whereas the second is perhaps to a god delegating to a world. An alternative pair of analogies is that of a correspondent writing a letter vs a novelist writing a novel. Oettinger's (1969) distinction between computers as actors and as instruments is isomorphic to the beings-worlds dichotomy used here.

The usefulness of the beings-worlds classification is discussed below, once again by referring to issues relating both to personality assessment and to human-computer interaction.

As Feigenbaum and McCorduck (1984) point out, the idea of artificial beings is not modern. Robots are described in classical Greek literature. Computer-like "brazen heads" that solve mathematical puzzles were thought to exist in the Middle Ages. An artificial human-like clay creature, the Golem, was at one time rumoured to stalk the streets of Prague. The Frankenstein monster is a more recent fictional example of a human-made creature (Shelley, 1978).

The idea of remote-controlled beings is also old. There is a very large folk literature on people and animals being put under spells and made to act the will of their enchanters. A rather sinister version of this idea that persists into the present is the cult of the zombies - animated corpses under the control of a malevolent corpsemaster.

These artificial and remote controlled "beasties" perhaps owe their strange attraction to the ambivalent feelings they seem to evince in us. On the one hand the cues prompt us to act as if we are dealing with a sentient being, on the other we are only too horribly aware of the human creator/manipulator behind the scenes. On a more mundane level, this is the same kind of ambiguity we experience upon reading the words "WASH ME" traced in the dust on a dirty truck. As Hofstadter (1979) explains, "one is supposed to pretend, on some level, that the truck itself wrote the phrase and is requesting a wash. On another level, one clearly recognizes the writing as that of a child, and enjoys the humor of the misdirection" (p. 608). For the child (or shall we say the test constructor) there is also an ambivalent thrill - what was intended as a joke, suddenly seems to rear up with a life of its own.

Given the long history of artificial beings in the folk consciousness, it is not surprising that computers would be cast in this role. The anthropomorphized computer has become a stock character in science-fiction movies, vying with human characters for prominence in classics of the genre such as "2001: A Space Odyssey" and its successor "2010".

In more intellectual circles the tendency to view computers as potential artificial beings has been no less strong. The most widely agreed on test for computer intelligence, the Turing test, involves a human conversing with either a computer or another human via a terminal and then having to judge whether his or her conversation partner was human or machine (Prenis, 1981). As soon as human judges can no longer reliably make this distinction computers can be said to be exhibiting intelligence. The idea that machines could only be said to be intelligent if they could converse like humans is very old, having already been suggested by Descartes in the 16th century (Regan & Singer, 1976). The most successful branch of AI, that dealing with "expert systems", can be said to represent an attempt to pass the Turing test in limited domains in that it attempts to emulate the actions of human experts (Michie, 1982).

Computers as artificial beings is a potent image, then, both in the public and in the scientific imagination.
Computers-as-worlds, on the other hand (remembering the beings-worlds dichotomy), is not such a strong image, despite the fact that the majority of computer programmes nevertheless probably function as artificial worlds rather than as artificial beings. Statistical analysis programmes, data base programmes, word processing programmes, spreadsheet programmes - all can be said to constitute small artificial worlds. When such programmes function optimally, the computer as such fades into the background, becoming a mere facilitator of the user's interaction with the world of words, figures, personnel records, or

whatever. Oettinger (1969) labels this 'fading into the background' of computers in their role as instruments (i.e. as worlds) 'transparency'. The relative transparency of the computer is an important characteristic of artificial world programmes. When a programme presents a user with an "artificial being" situation on the other hand, it is only natural for the user to see the computer itself as constituting the being.

Computer games provide an illustration of the way the worlds-beings dichotomy relates to the computer's transparency. Ever since the earliest narrative "adventure" games were created, computer games have overwhelmingly been of the artificial worlds type. Whether the action is verbally described or visually depicted and whether the player's aim is to find the hidden treasure; "zap the aliens"; or "gobble up the blobs" - the interaction is with an artificial world rather than with an artificial being.

Perhaps closest to a beings type game is computer chess where the computer takes on the role usually occupied by a human opponent. It is perhaps for this reason that chess is the only computer game to be drawn into the machine intelligence controversy, with human vs computer tournaments being anxiously watched for signs that computers may be rivalling expert human players (Dreyfus & Dreyfus, 1986).

In theory, then, computer games are mostly artificial worlds and the computers administering the games mostly transparent. However in practice the computer is often <u>not</u> transparent. This

non-transparency is due to what may be termed intrusions by the computer-as-being into the game-world. The first type of intrusion is a consequence of actions by the human player that fall outside the scope of the rules of the game. In a narrative adventure game, for instance, the player may type in something like: "Invert the table" to which the computer is likely to respond: "I don't know what 'invert' means." In more serious applications this type of intrusion occurs when the human user attempts something not foreseen by the programme, resulting in requests for clarification, error messages, or even programme failure - all intrusions that draw attention away from the application world and onto the computer as (recalcitrant) being.

Another class of intrusions that occur in computer games are deliberately programmed messages from 'the computer', often in the form of comments on the player's progress or lack thereof. This kind of intrusion is probably, consciously or otherwise, intended to invoke in the player the mythical awareness of artificial beings discussed above, and thereby to enrich the game. Depending on the player's feelings about the mythology surrounding artificial beings (or perhaps even about the artificial intelligence controversy) he or she may find such intrusions funny, exciting, inspiring, annoying, or childish. In Chapter 5 the avoidance of such intrusions will be demonstrated in an actual test.

The practical application of the beings-worlds dichotomy is discussed below.

A computerized test where the assessor has delegated to an artificial being is an attractive prospect and artificial beings scenarios come most naturally to mind when thinking about personality assessment by computer. So tempting is the image of the computer as an artificial interviewing psychologist that questionnaires presented on computer routinely have their status enhanced to that of 'interviews'. Erdman et al. (1981) cite more than twenty references to such "interviews". As was discussed in Chapter 2, the potential for interview-like flexibility is the one aspect of computerized personality assessment that has most captured the imagination of writers in the field.

Despite the intuitive appeal of artificial beings tests on computer, the drawbacks are severe. When the delegation is to a being rather than to a world, the computer itself, as has been discussed, becomes salient. This means that the computer itself becomes a social stimulus affecting the testee's responses. If attitudes towards computers had become sufficiently crystallized in the testee population, the impact of the computer would be more or less standard for everybody. However it is the author's contention that there is as yet no consensus as to how to respond to computer's emotionally; or how to fit them into our mental representations of the world. What kind of beings <u>are</u> computers - Frankenstein monsters, zombies, or perhaps nothing more than benignly passive brazen heads? As was discussed in Chapter 2, even a comparatively simple question such as whether computers are likely to elicit more or less honest responses cannot be answered because cultural attitudes towards computers are still evolving.

At present, the more cautious approach would therefore seem to be to construct interactive personality tests on computer as imaginary worlds tests, taking care to avoid intrusions by the computer-as-being into the test world.

In summary, it has been suggested that the second way in which the mediational function of tests can be classified is in terms of delegation. The assessor can delegate either to an artificial being or to an artificial world. Although artificial beings are more prominent in the scientific and folk imagination than artificial worlds and computers are often seen as artificial beings, computers have in practice been used much more frequently in the artificial worlds sense. In artificial beings applications the computer becomes salient, whereas in artificial worlds applications the computer becomes transparent. Given the fact that attitudes towards computers are still in flux, the more cautious approach at present is to construct interactive personality tests as artificial worlds rather than as artificial beings.

3.3.3 Subtlety (questions vs koans)

The third dimension included in the model is that of subtlety. The question as to how "obvious" or "subtle" personality tests should be has in one guise or another been investigated in vast numbers of research projects. At the widest level, the sniping that still occurs between the "projective" and "objective"

schools is based partly on differing stances regarding this issue: The "projectives" pride themselves on the subtlety of their instruments, while the "objectives" scoff at their apparent

straightforward or obvious, and the interrogator's intent and the meaning of the respondent's responses are relatively clear. Koans are more subtle. The word is borrowed from the Zen-Buddhist term for the interrogative puzzles set by masters for their pupils (Kopp, 1978). A well-known example is: "What is the sound of one hand clapping?" A pupil would typically spend days or weeks pondering the koan, sometimes submitting laboriously constructed solutions only to find them rejected. Then one day enlightenment strikes and the pupil submits the correct answer (often as cryptic as the question); the master smiles enigmatically, and serves up another koan. Thus koans resemble questions in that they require a response, but differ from questions in that neither the interrogator's intent nor the meaning of the response is clear to an outsider. An item from the CPI (Gough, 1956) could be classified as a question; an inkblot as a koan.

The question-koan or subtle-obvious dichotomy has been treated at length and from many different angles by other authors, as was indicated in the introduction to this section. As has been discussed, opinions differ sharply on the relative merits of subtlety and obviousness. There is no need to enter the fray here. Rather an effort is made to define how the dichotomy manifests itself when analysed in terms of the assessment technology used.

It will be recalled that assessment technology plays a mediational role between the assessor and the assessee. For conventional tests, a rather elegant method of defining subtlety

in terms of the mediated interaction between assessor and assessee has been developed by Holden and Jackson (1985). Holden and Jackson explain subtlety in terms of the interaction between "face validity" and the actual scales comprising a test. Face validity is "the contextual relevance of a test item for a respondent" (p. 217) and is operationally defined as the ease with which testees are able to group a particular item with other items in what they perceive to be a scale. Subtlety is inversely related to the degree that testees' groupings coincide with the actual scales used by the tester in scoring the test.

An interactional approach such as Holden and Jackson's (1985) suggests that subtlety or obviousness is not an objective feature of a test item, but is determined jointly by the assessor's and the assessee's experience of the item. In a conventional test this joint experience occurs in what one may term a "tenuously mediated" form: Judgements about the meaning of a particular item or response are made at different times and without any opportunity for mutual consultation. In interactive tests the situation is different: The interaction is more "closely mediated" as at least some consultation is possible about how items and responses are to be interpreted. In other words, it is possible to negotiate about how the unfolding situation is to be interpreted - irrespective of whether the negotiation is amicable or takes the form of a battle of wills.

Negotiating about the interpretation of reality rarely happens explicitly in real life, of course - except in special circumstances such as in court cases. That this kind of

negotiation nevertheless does occur (as the "ground" against which the contents of what we do and say is drawn as "figure") is a central tenet of systems and interpersonal approaches in psychology (Watzlawick, Weakland & Fisch, 1974). Non-verbal communication typically provides some of the "ground" in interpersonal interaction. Natale (1975) reviews an impressive collection of studies showing how partners in an interaction tend to move towards agreement, as measured in the "convergence" of such non-verbal indices as speech latency, speech rates, and vocal intensity. Natale found that a person's relative willingness to converge (i.e. his relative pliancy) correlates with his or her tendency to respond in a socially desirable direction to personality questionnaire items; this provides more evidence in support of the opinion (held, for instance, by Jackson & Helmes, 1979) that social desirability responding may be predictively useful.

However the negotiation which occurs as 'ground' to whatever else happens is not limited to non-verbal communication. Each utterance a person makes in an interpersonal setting has at least two levels of meaning: The overt content level (i.e. the figure) and the covert relationship level (i.e. the ground). The figure is typically in focus while the ground is out of focus. The ground or covert meaning can be brought into focus by attending to non-verbal cues but also to the context in which an utterance occurs. Kiesler's (1983) system, to be discussed in Chapter 5, is an example of how this can be done. Depending on what preceded it, an apparently hostile statement such as "You shouldn't have done that" could, using Kiesler's taxonomy, be

decoded as an attempt to keep the relationship from disintegrating. Other systems for deducing the covert meaning from the context of the interaction also exist. A considerable amount of research has for instance been conducted on reconstructing the tussle for dominance that goes on in a conversation from the pattern of topic changing and topic following of the interactants (see Friedlander & Phillips, 1984; Tracey, 1985; Tracey & Ray, 1984). Contextual cues as to the covert meaning of what a person says or does can be monitored in interactive tests as readily as in real-life conversations.

Which then is the most promising way of constructing personality tests on computer - as questions or as koans? From the considerations outlined above the answer would seem to be - as questions. At least three reasons can be found for this: One, the social desirability responding that goes with relatively obvious tests appears to be predictively useful (Angleitner & Wiggens, 1986). The apparent fakability of such tests can be said to mirror the fakability of real life. Two, for some sort of negotiation to be possible about how the test situation is to be interpreted there must be the potential for agreement; at least to this extent the test must be obvious. Three, assessing a person in an interactive situation makes it possible to pick up cues as to the covert meaning of his or her actions from the context of the unfolding interaction. The ability to fake on the ostensibly primary (or "in-focus") content level does not necessarily imply an ability to fake on the "out-of-focus" relationship level.

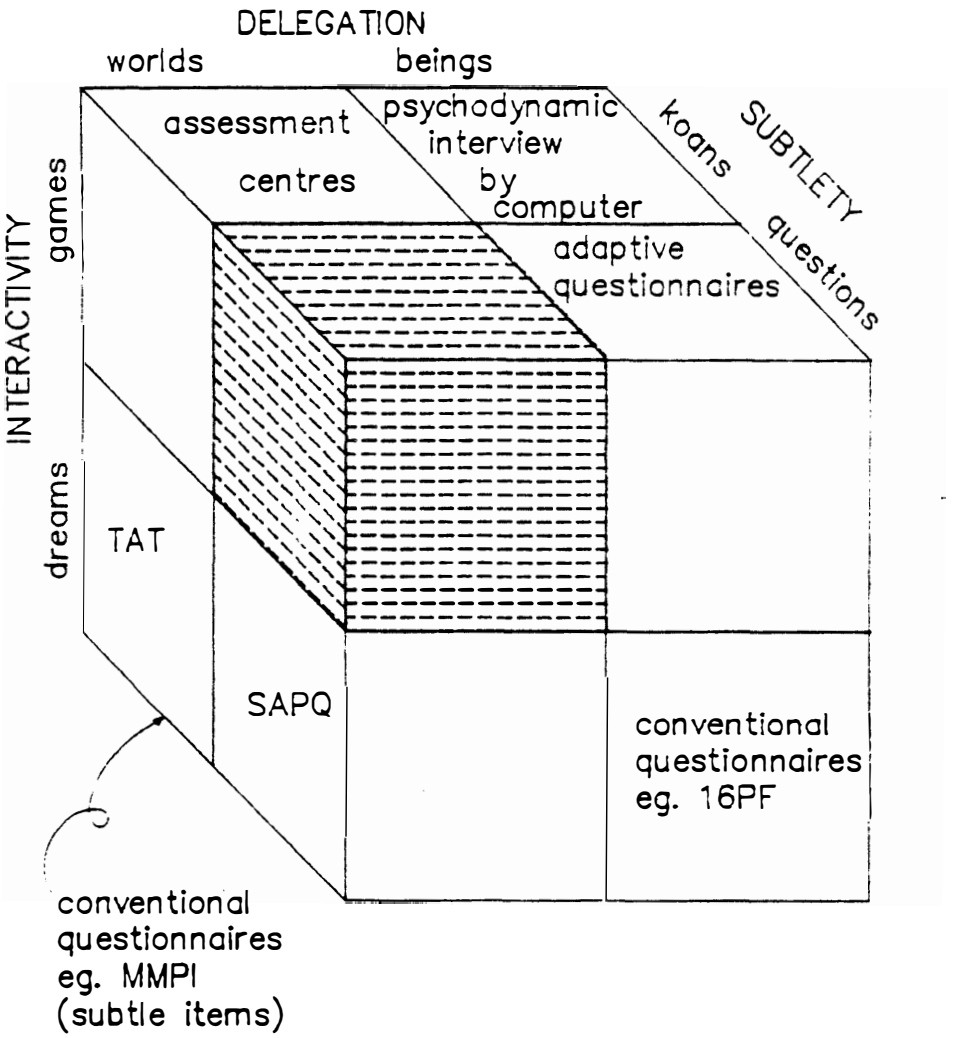
In summary, it has been suggested that the subtlety-obviousness dichotomy is one of the most important in personality assessment and therefore merits inclusion in the model. Looked at from the point of view of the assessment technology used, it becomes clear that subtlety is not an objective property of an item, but is determined jointly by the assessor and assessee. In an interactive test this takes the form of an out-of-focus negotiation about the interpretation of the unfolding situation. Since the obviousness of conventional questionnaires seem to add to rather than detract from their validity, since a degree of obviousness is needed to make the simulation of interpersonal negotiation possible, and since interactive tests in any case make the assessment of covert behaviour possible, it can be concluded that a fair degree of obviousness is desirable in an interactive personality test.

3.4 Conclusion

The model developed in this chapter is illustrated in Figure 3.1. An example of an assessment procedure is given for each cell in an effort to illustrate the range and something of the flavour of the model. To use computers optimally for personality assessment, the shaded-in cell is judged the most promising of the eight cells in the model. Tests in this cell are what were termed "phase three" tests in the previous chapter and can be characterized as "games", "worlds" and "questions". That computerized tests should be constructed as "games" (i.e. interactively) if full use is to be made of the potential for

innovation that computers offer, seems virtually certain. The case for constructing computerized tests as "worlds" and as "questions" has less intuitive appeal, although arguments have been advanced to show that this could be the best route.

Figure 3.1: A model of assessment as a relationship mediated by technology.



Looking at the three cells bordering directly on the "phase three" cell (i.e. cells that fall in the same category as the shaded-in cell on two of the three dimensions), it is interesting

to note that none of them contain conventional personality questionnaires such as the 16PF or CPI. Conventional personality questionnaires share only their relative obviousness with the "phase three" tests proposed in this report. Tests that fall in cells bordering on the "phase three" cell are adaptive questionnaires ("phase two" personality tests), assessment centers, and the SAPQ (Steyn, 1977). That "phase two" tests should be related to 'phase three' tests is not surprising. The fact that assessment centres and the SAPQ are related to 'phase three' tests suggests that "phase three" tests may in some sense be viewed as simulations. Assessment centers are explicitly designed as simulations of reality. The SAPQ can be viewed as a simulation to the extent that its items consist of descriptions of situations and possible responses to each situation, rather than of self-descriptive statements as in for instance the 16PF. Assessment centers and the SAPQ may also be useful in determining the construct validity of early "phase three" tests.

4. IMPLICATIONS OF INTERACTIVE ASSESSMENT

In the previous chapter it was decided that the crucial difference made by the introduction of computerized testing is that interactive assessment becomes possible. If "interactive" is to be more than just a label for any computer-administered test, however, one needs to define the term more closely and to explicate how it impacts on psychological and psychometric theory.

The Concise Oxford Dictionary defines interaction as acting reciprocally or on each other. Thus an element of reciprocality is required for a test to be accurately termed interactive - the testee's actions must be contingent on the test's actions and the test's actions must be contingent on the testee's actions.

The possible advent of new personality testing techniques are likely to be due (as discussed in the first chapter) to advances in computer technology and not to advances in psychometrics or personality theory. It is therefore to be expected that conventional psychometrics and personality theory will show signs of strain when applied to the new kinds of testing. The nature of these strains and possible ways of overcoming them are discussed in the next two sections.

4.1 Psychometrics and interaction

Traditional psychometrics is unquestionably not geared to dealing with interactive assessment. The stumbling block is an assumption that underlies both classical test theory and the more recent item response theories, namely that items are "locally independent" (Coulter, 1973; Morf, Alexander & Fuerth, 1981). Local independence means that responding in a particular manner to an item (or even the fact of having read the item) should have no effect on responses to later items. The local independence assumption can also be termed an assumption of no context effects (i.e. the context or position of an item in a test does not affect a testee's responses) or an assumption that observations are independent (i.e. the "observation", made by means of the test, of the testee at item 1 does not affect the "observation" at item 10). These terms are used interchangeably in what follows.

The independence assumption is clearly quite contrary to what is implied by the very notion of interactive testing. If, as was decided above, interactive assessment implies reciprocal causality between the test and testee, then observations cannot be other than dependent since the nature of the observer (the test) is changed by what is observed.

The independence assumption applies to many statistical procedures, not only those used in psychometrics, and can be traced back to the development of analysis of variance for early agricultural research - a field in which interdependence of

observations is unlikely to be a significant factor (Kenny & LaVoie, 1985). Psychological statistics in many cases still operate in the old Cartesian paradigm which deals in unilateral causality and assumes that context-free "readings" of natural phenomena can be made (Dreyfus, 1979). The modern trend, evident in physics (Capra, 1976) as much as in psychology (Silverman, 1977), of including a description of the observer in a description of the observed, has not yet made an appearance in psychometric theory.

In the field of cognitive testing there has been occasion to question the independence assumption, particularly as it relates to fatigue and practice effects (Kingston & Dorans, 1984; Wing, 1980), "context effects" in adaptive tests (Kingston & Dorans, 1984) and the effects of providing testees with feedback (Betz, 1977). In personality testing, also, the assumption appears at times to be violated. This is discussed below.

4.1.1 Conventional personality tests

The assumption that items are locally independent is perhaps even more difficult to accept in personality testing than in cognitive testing. In cognitive tests it is still possible to imagine a testee working on each problem independently, but in personality testing it seems not unlikely that testees would consciously relate items to each other - evolving a strategy of responding as they proceed through the test. A conscious drive to be consistent or a counterbalancing style are examples of strategies

testees may follow. The literature on such context effects is reviewed below. The intention is to demonstrate the extent to which the independence assumption is violated, even in conventional, non-interactive tests.

The development of MMPI (Hathaway & McKinley, 1951) "short forms" gave the initial impetus to research on item context effects in personality questionnaires. The concern was that MMPI items taken out of their original context and regrouped in a short form might function differently than in the original test. Two early studies (Perkins & Goldberg, 1964; and Weiss & Moos, 1965) both concluded that response interdependence is not a significant factor in the MMPI. However the generalizibility of these findings can be questioned. The MMPI item pool is notoriously heterogeneous and many items are of the so-called "subtle" type. It is therefore perhaps not surprising that testees find it difficult to relate items to each other and instead tend to respond to items independently. Subsets of particularly homogeneous items in the MMPI might provide different results. In a study using only the 16 repeated items in the full-length MMPI (a situation involving maximal homogeneity between pairs of items), McGrath, O'Mally and Dura (1986) found contextual effects: There was a consistent tendency to respond in a less pathological direction on the second presentation of each item.

That contextual effects could be a factor in more homogeneous questionnaires is illustrated by two studies (Kuncel, 1973; Tuck, 1982) where such effects were found in, respectively, the Personality Research Form and the Achievement Anxiety Test.

Solomon and Kopelman (1984) found that the reliability of a questionnaire depends on whether items are grouped together in scales or not; grouped scales were more reliable than scales interspersed with items from other scales. In a similar vein Schubert and Fiske (1973) found that presenting the MMPI repeatedly increases the reliability of each successive presentation. In the field of opinion surveying it is an accepted fact that the order in which questions are presented can have an effect on responses (Sudman & Bradburn, 1983).

Between-tests contextual effects have also been found in the literature. Clayes, De Boeck, Van den Bosch, Biesmans and Bohrer (1985) found that preceding a personality questionnaire with a self description exercise almost invariably increases the questionnaire's predictive validity. Similarly, Osberg (1985) found that the order in which two personality questionnaires are administered affects their respective validities.

In summary: There is clear evidence to suggest that the independence assumption is violated in conventional personality questionnaires. Only the extent to which this occurs is a matter for debate.

4.1.2 Interactive personality tests

Item interdependence is perhaps at most an irritation in conventional tests, but it could present a serious problem in interactive tests. In an interactive test each individual is

typically presented with a unique subset of items. If one cannot assume that items function independently of each other, it becomes impossible to obtain summary response scores on which testees may be compared since no two items are ever exactly alike. This problem has been discussed at some length by Kingston and Dorans (1984). They found that contextual effects do play a role in adaptive cognitive tests. No research on context effects in adaptive personality tests could be found, however.

4.1.3 Possible solutions

Given that conventional psychometrics is not geared to deal with interaction, what are the possible solutions? Two classes of solution are discussed below - statistical and pragmatic solutions.

Statistical solutions involve improving currently used psychometric statistics or introducing statistical methods better suited to the interactional context. Kingston and Dorans (1984) recommend the improvement of current item response models to incorporate parameters related to item interdependence. A statistical model of this sort has recently been developed for certain cognitive testing situations by Andrich (1985). Improved item response models of this kind are however at best an attempt to improve the efficiency of measurement of conventional

personality constructs and do not attempt to focus on response interdependence as an individual difference variable in its own right.

There are statistical methods which focus directly on response dependencies. These are called process statistics and include methods such as Markov chain analysis, lag sequential analysis and uncertainty analysis (Greenberg, 1986). Process statistics can be used to quantify the sequential dependencies in interactional settings such as interviews (e.g. Anchin, 1982; Duncan et al., 1984; Friedlander & Phillips, 1984; Russel & Trull, 1986; Tracey & Ray, 1984). Smiling behaviour in an interview can be used to illustrate the inadequacy of using conventional statistics in this kind of setting. Simply counting the number of times an interviewee smiles, as would be done when conventional statistical methods are used, could result in two interviewees with quite different smiling styles getting the same 'smile score'. The first interviewee might smile mainly in response to the interviewer smiling whereas the second interviewee might initiate smiling without cues from the interviewer (Duncan et al., 1984). These kinds of sequential dependencies can be captured by means of process statistics. Conceptually, the difference between conventional and process statistics is a matter of representation. The former represents the behaviour of some actor as if it occurred unilaterally, whereas the second includes a representation of the context (i.e. the observer) in its representation of the actor's behaviour (Keeny & Morris, 1985).

To the author's knowledge process statistics have not been used to measure individual differences in test responses. As computerized interactive testing gains ground, however, this may well happen. Reciprocal causality, which is here taken as the defining characteristic of interactive tests, demands that in one way or another a representation of the test's actions be included in any description of the testee's actions. Should this be done by means of process statistics poetical justice would be done, since it was to describe the operation of early computers that process statistics were first developed (Sackman, 1967).

In summary, two classes of possible statistical solutions to the 'local dependence' problem are adapting present item response models to incorporate parameters related to interdependence, or using process statistics. The former has the advantage that the development of such models is already well under way, the latter that it can be used to represent behaviour in an interactional setting more effectively.

Pragmatic solutions involve designing interactive tests in such a way that the local independence problem is minimized.

Kingston and Dorans (1984) mention two possible pragmatic solutions - giving all testees sufficient practice and dropping those items from a test that are found to be particularly context-susceptible. The first of these two options is only applicable to cognitive tests, while the second might involve throwing out valid variance together with the offending items.

Another pragmatic solution would be to limit the extent to which an interactive test can "branch" depending on a testee's responses. If only a fairly small amount of branching is permitted, testees responses would at least be comparable to those of other testees "on the same branch". Again the disadvantage is that valid variance may be lost in direct proportion to the degree that interaction or branching is limited.

A fourth pragmatic solution would be to appeal to the face validity of an interactive test. If such a test could in some sense be said to simulate real life, then the particular branchings a testee follows through the test (i.e. the particular way in which the test acts upon the testee) would be similar to what occurs in real life and the interdependence of items would simply reflect real-life interdependence. More proof is needed, of course, to substantiate a claim that a test simulates real life than is needed to claim (as for conventional tests) that test responses co-vary with real-life behaviour.

Finally, one could refuse entirely to be put into the psychometric straightjacket and design interactive assessment instruments without regard to traditional psychometric concerns. However it may be well not to forget the "prosaic but important point" made by Hunt & Pellegrino (1984, p. 34) that problems of measurement rigour do not go away simply because a computer has been introduced into the assessment situation. Experience with MAZE, a computerized management simulation designed by Tredoux (1985), illustrates the point nicely. The simulation presents

the testee with a rich variety of simulated situations and makes full use of the computer's branching capabilities, but scores have proved difficult to analyse psychometrically.

In summary, pragmatic solutions to the local independence problem appear to involve having to walk a tightrope between losing valid interactional variance and losing measurement rigor.

4.2 Personality theory and interaction

The fact that interactive assessment involves a degree of interpersonal game playing (as discussed in the previous chapter) suggests that game playing skill could become a factor in testees' performance. Skill is of course traditionally thought of as a cognitive rather than a personality attribute. In terms of Cronbach's (1984) well-known classification of tests of maximum vs. tests of typical performance, personality assessment has traditionally been seen as belonging to the typical performance class. The term "personality" is ambiguous, however. In its broadest sense it refers to all enduring non-physical characteristics of a person, including such things as intelligence and aptitude. It has frequently been suggested that in accordance with this definition the term should not be limited to non-cognitive factors. Taylor (1987), for instance, thinks that concepts at the personality-cognition interface (such as impulsivity and rigidity) could prove useful in future test construction efforts involving computerized administration. The more usual definition of personality as "emotional, motivational,

interpersonal, and attitudinal characteristics, as distinguished from abilities" (Anastasi, 1982, p. 497) is however adhered to in this report.

Personality theories of the non-cognitive sort can be grouped into four classes: trait, psychodynamic, situational and interactional (Kunce & Kunce, 1982). Each class of theory has a particular stance with regard to the trait-situation controversy of behavioural causality. At the one extreme are classical trait and psychodynamic theories that view behaviour as being virtually exclusively shaped by inner forces; at the other extreme is situationism (i.e. classical behaviourism) that views behaviour as exclusively externally determined. Interactional theories represent a compromise position between these two extremes. Since interactive personality tests have been defined as being characterized by reciprocal causality it seems unlikely that theories that unilaterally emphasize either inner or outer causality would be suitable for interactive testing. Interactional theories, on the other hand, by their very name suggest that they may be suitable for interactive testing. Two kinds of interactional theories can be distinguished, here termed situational interactionism and interpersonal interactionism.

4.2.1 Situational interactionism

Historically, many theories of personality, including those which were developed a considerable time ago, can be classified as theories of situational interaction. Examples include Murray's

need theory (Murray, 1938, 1962), which seeks to describe behaviour in terms of the interaction (themas) relating personal needs and environmental press; Lewin's (1935) dynamic theory of personality, which is based on the interactionist position described by the famous equation B=f(P,E) (behaviour is a function of the person and the environment); and ecological psychology (Barker, 1968) which, inter alia, attempts to characterize human lives in terms of particular "behaviour settings".

A more recent situational interactionist theory can be loosely termed "modern interactionism" (Epstein, 1979). Modern interactionism grew from dissatisfaction with the one-sided, static trait theories typically associated with personality questionnaires. In particular, Mischel's (1968) harsh criticisms of "the assumption of massive behavioural similarity across diverse situations" (p. 296) led to attempts to find ways of describing personality in terms of the interplay between persons and situations. Mischel's own theory, cognitive social learning theory, is one such an attempt.

Much of the research that has been conducted within the modern interactionist paradigm has been inspired by Endler and Hunt's (1966, 1969) work with the S-R inventory of anxiousness. Endler and Hunt argue that conventional personality questionnaires one-sidedly emphasize the response side of behaviour, while the situation side is limited to descriptions such as "always", "sometimes" and "never" (Endler, Hunt & Rosenstein, 1962). In the S-R inventory the testee is presented with a number of

descriptions of anxiety-provoking situations (e.g. "giving a speech"; "being trapped in a lift"). For each situation the testee has to indicate his likely responses from a standard list (e.g. "heart beats faster"; "feel scared"). Analysis of variance of S-R inventory scores indicated that both persons and situations contribute little to the total variance, but that the interactions between the two contribute a substantial proportion. This suggests that people don't have strong behavioural predispositions that apply over all situations, nor do particular situations elicit similar responses from everybody, but that there are constancies in the ways particular people react in particular kinds of situations.

The Endler and Hunt (1966, 1969) studies have generated a great deal of research controversy. The issues in dispute include the correct way of applying the analysis of variance technique to test for person x situation interaction (see Ayton & Wright, 1985; Furnham & Jaspers, 1985) and the extent to which situations and response modes have been adequately sampled (Cartwright, 1975). Despite these controversies, the idea of obtaining information on the person-in-a-situation, rather than just on persons or just on situations, seems a good one. With person-in-situation information, as Sundberg (1977) points out, it would be possible to "make conditional predictions such as if individual A is in situation M, he is likely to do Z; if A is in situation N, he will probably do Y; if B is in situation N, she will probably do W" (p. 130).

If person-situation interactions are to become the unit of measurement, taxonomies of situations to complement trait taxonomies have to be found. Various such taxonomies have been proposed, e.g. "behaviour settings", "event structures", "games", "scripts" and "frames" (Minsky, 1974), but as Sundberg (1977) points out, "the problem is the absence of ways to classify situations and individual difference variables in a congruent manner" (p. 286). This author draws attention to the problem that situational taxonomies are typically vast; however, only a short taxonomy of situations related to a small taxonomy of traits "will lead to an enormous number of interactions" (p. 286).

Another problem with the situational interactionist position is that it tends to assume that persons and situations have a unidirectional influence on responses; consequently it ignores the possibility that responses might in turn shape situations and persons. As Emmons, Diener and Larsen (1985) point out, the "interaction" in modern interactionism refers to the statistical interaction between variables and not to what might be termed reciprocal, dynamic, transactional or organismic interaction where "situations and persons are at the same time both independent and dependent variables, and there is bidirectional causality between the two" (p. 693). Even the limited interactionist possibility that persons may consciously <u>choose</u> situations (Gorta, 1985; Magaro, Ashbrook, Lesowitz & Johnson, 1985) is not catered for by statistical interactionist models of the Endler and Hunt type.

In summary, it would appear that although the situational interactionist position holds promise for use in interactive personality assessment, the incongruence between taxonomies of persons and of situations and the failure to accommodate true bidirectional interaction present problems.

4.2.2 Interpersonal interactionism

Like situational interactionism, interpersonal interactionism seeks to describe personality in terms of both persons and situations. The difference is that interpersonal situations are seen as being of primary interest. Other people are seen as constituting an individual's situation, thus at once overcoming the trait-situation taxonomic incongruence problem described above. The price that has to be paid for this is in narrowing the scope of the term "personality". Rationalizations usually advanced by interpersonal theorists in this regard are that interpersonal situations constitute the most important class of situations and that "even impersonal situations have interpersonal components in the form of imagined or otherwise symbolized presences" (Kiesler, 1982, p. 5).

Interpersonal theories of personality are currently enjoying considerable popularity in psychology. This can be seen as a manifestation of a historic trend in personality theorizing away from exclusively intrapersonal and towards more interpersonal conceptualizations. Leary (1957) points out that virtually all major personality theorists since Freud have made some attempt to

include interpersonal aspects in their theories. Some of the examples listed by Leary are Horney (who emphasized cultural influences), Fromm (who saw personality as the product of human relations rather than instinctual pressures), Erikson (who 'socialized' Freud's psychosexual stages) and Sullivan (who saw personality as an explicitly interpersonal phenomenon).

Sullivan's (1953) definition of personality is still accepted by many interpersonal theorists as the best available. According to the definition, personality is "the relatively enduring pattern of recurrent interpersonal situations which characterize a human life" (p. 110 - 111). Sullivan described interpersonal situations as forming an 'enduring pattern' because they are in some measure 'set up' by a person. Some people, for instance, repeatedly find themselves in the position of being exploited by others - this can be assumed to be due, at least to some extent, to the fact that they seek out and create situations in which they function as victims. Sullivan (1954) brought the important concept of participant observation into Psychiatry. Participant observation constitutes a recognition of the fact that the social scientist is part of what he is observing; "independent" observations are impossible.

Sullivan's theorizing was in many respects very general, and Leary (1957) set out to define some of his concepts in greater detail. His "interpersonal circle" is an attempt to describe the ways in which people create the situations that typify their lives. The interpersonal circle is at the same time a taxonomy of interpersonal styles and of interpersonal situations (since

other persons form an individual's situation) . Behaviour in any particular segment of the circle constitutes a "bid" for the other person to act in the "complementary" segment of the circle. Bids are made so as to push others to confirm one's self-definition. This is done to avoid the anxiety of having one's self-definition disconfirmed.

Elaborations on Leary's (1957) interpersonal circle have been proposed by, among others, Wiggens (1979), Conte and Plutchik (1981), Buss and Craik (1983), and Kiesler (1983). A-priori criticisms such as that personality "is almost certainly more complex than which can be represented realistically in a two-dimensional plane" (Jackson & Helmes, 1979, p. 2284) have been advanced against the interpersonal circle. That the circle offers an elegant solution to the trait-situation incongruence problem cannot be argued, however.

Apart from Sullivan's (1953, 1954) and Leary's (1957) work, another important input to modern interpersonal theory has been systems theory. Systems theory is an interdisciplinary scientific approach that has found application in fields as diverse as Chemistry, Biology, Sociology and Psychology (Miller, 1978). The theory is concerned with how elements are configured in wholes rather than with linear causality (Emery, 1981). It proposes a hierarchy of systems, where every lower level system is an element, or subsystem, in a higher level system (Jordaan & Jordaan, 1980).

It is theoretically possible to focus on a person as a system composed of subsystems, i.e. intrapersonally, as has been done by Royce (1979) and Angyal (1981), but the majority of systems theorists in psychology have preferred to focus on the person as an element in the higher-level systems of interpersonal relationships (e.g. Keeny, 1979; Searight & Openlander, 1984). Systems theorists thus typically see pathology as being seated in the group rather than in the individual; for instance a person may behave in a schizophrenic fashion because of his function in a pathological family system.

Because systems theorists and interpersonal theorists both emphasize the relationships among people, their work overlaps to a large degree. An extreme systems position is to deny the existence of personality altogether, ascribing all behaviour to the functioning of the group, but in practice most systems theorists would concede that an individual has a personality to the extent that he creates "similar systems out of all his relationships - in other words the relationships in which he finds himself are to a large degree defined by his peculiar style" (Cook, 1984, p. 6).

It is clear that interpersonal interactionism, unlike situational interactionism, explicitly caters for bidirectional causality between persons and situations. Thus it seems likely that it could provide an appropriate theoretical base for interactive personality assessment. In the next chapter a particular model

of interpersonal behaviour, Kiesler's (1983) "taxonomy for human complementarity", is discussed. Kiesler's model will be used as a guide to the construction of an interactive test.

4.2.3 Conclusion

In this chapter the implications of interactive assessment for psychometrics and personality theory were discussed. It was concluded that mainstream psychometrics is not suitable for use in interactive assessment and that statistical or pragmatic alternatives would have to be found. With regard to personality theory it appears that only interactionist theories, and in particular interpersonal interactionist theories, would be equal to the task of representing interactive assessment.

5. THE CONSTRUCTION OF AN INTERACTIVE PERSONALITY TEST

In this chapter the practical application of the ideas and theories outlined in the previous chapters is demonstrated. The construction of the Relational Adaptation Profile (RAP), a personality test designed to make fuller use of computer technology than has thus far been the case, is described.

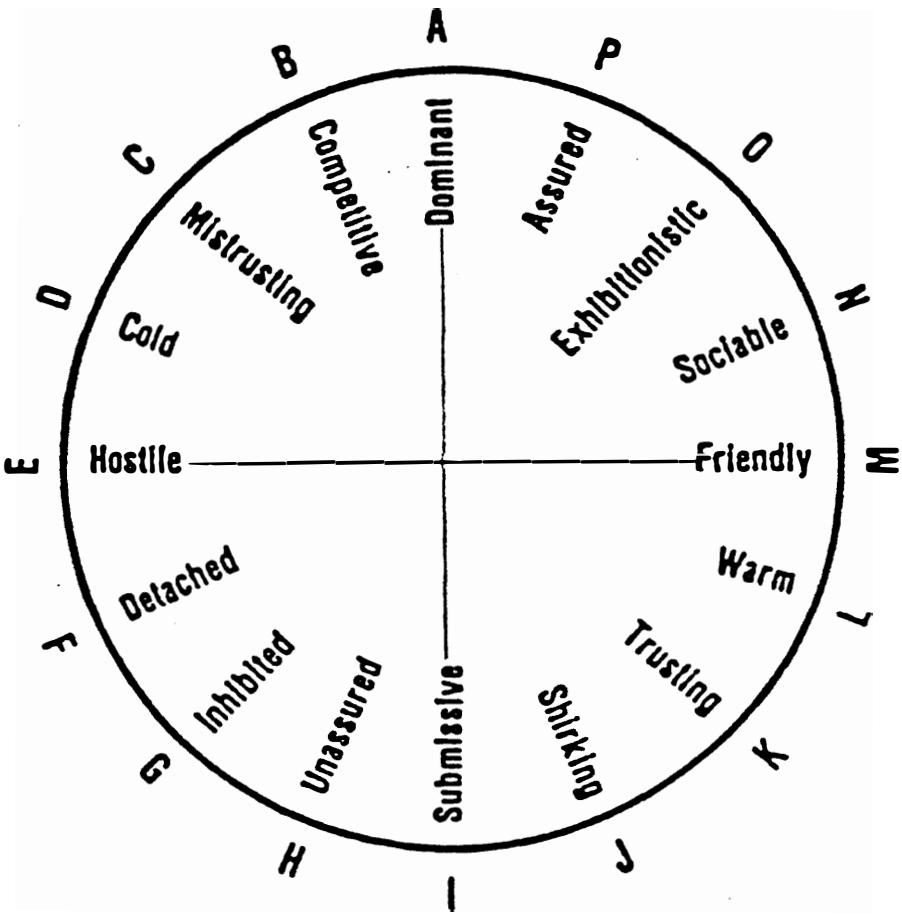
The specifications (decided upon in the previous chapters) for a personality test that makes full use of computer technology are the following. The first and most important requirement is that the test should be interactive, i.e. there should be reciprocal causality between the actions of the test and the testee. In order to deal psychometrically with the reciprocal causality statistical solutions (such as process statistics) or pragmatic solutions (such as limiting the amount of branching) or both will have to be instituted. To deal theoretically with reciprocal causality, the testee's actions will have to be described in terms of interpersonal theory. Finally, the test should be relatively unsubtle and should be presented as an imaginary world rather than as an imaginary being.

The construction of the RAP in accordance with these guidelines is described below in terms of the theory used, the development of a "script", the writing of items, and the development of scoring methods.

5.1 Theory

It was decided in the previous chapter that interpersonal theory is more applicable to interactive personality assessment than trait, psychodynamic or situationist approaches. Kiesler's (1983,1985) interpersonal circle (see Figure 5.1) is the most recent and comprehensive articulation of interpersonal theory and it was accordingly decided to use this as the basis for the RAP.

Figure 5.1: The Interpersonal Circle (adapted from Kiesler, 1985)

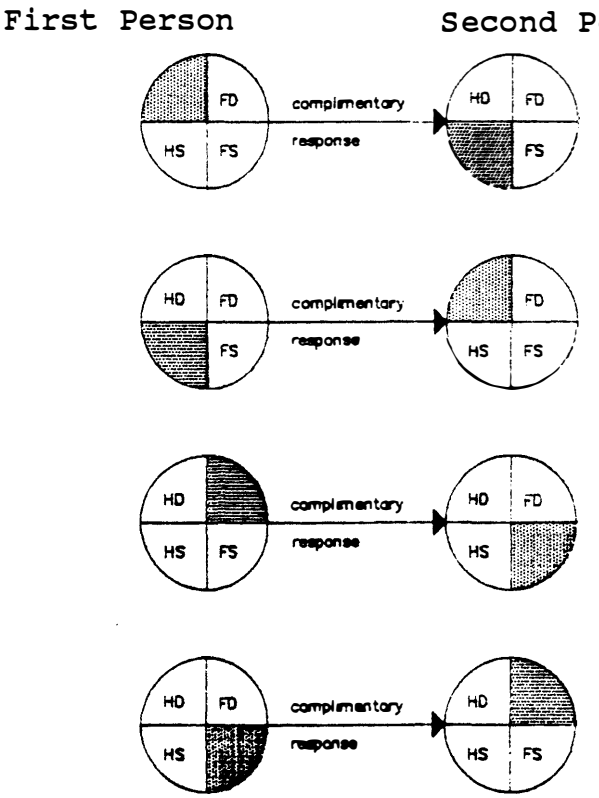


Kiesler's (1983 ,1985) interpersonal circle is a taxonomy of interpersonal behaviour in the form of a circle divided into quadrants by two intersecting axes labelled affiliation and control. The former is a hostile-friendly axis and the latter a dominant-submissive axis. The quadrants formed by the two axes

are: Hostile dominant (HD), friendly dominant (FD), friendly submissive (FS), and hostile submissive (HS). These can also be further subdivided into octants, and yet further into 16 circle segments.

Kiesler (1983,1985) provides detailed predictions regarding the kinds of behaviour that are likely to be elicited by actions drawn from each of the circle segments. The principle is that dominant behaviour is a bid for (is likely to elicit) submissive behaviour and vice versa, whereas friendly behaviour is a bid for friendly behaviour and hostile behaviour is a bid for hostile behaviour. Thus interpersonal behaviour can be seen as a bid to elicit reciprocal responses from others on the control dimension and corresponding responses on the affiliation dimension. Responses that can be interpreted as an acceptance of the other person's bids on both dimensions are termed "complementary" responses. Apart from complementary responses, Kiesler also identifies three other types of responses. The least likely to occur are anti-complementary responses. These are responses that reject the bids that the other person is making on both the control and the affiliation dimensions. An example would be responding with FS to another person's HS (to which the complement is HD). The remaining two kinds of responses are more likely than anticomplementary responses, but less so than complementary responses. They are termed semimorphic a-complementary responses (rejecting only the affiliation bid) and isomorphic a-complementary responses (rejecting only the control bid). The complementary relations between quadrants are illustrated in Figure 5.2.

Figure 5.2: Complementary responses



Note: Shaded-in areas indicate behaviour

Psychological ill health in terms of Kiesler's circle consists of rigidly persisting in behaviour from only a small number of segments and being impervious to other people's bids to get one to act in ways corresponding to other circle segments. The corollary to this is that psychologically ill people are very strong in bidding for particular kinds of behaviour from others and forcing them into acting in ways that confirm their self-definitions.

The interpersonal circle is rare amongst personality theories in that it provides a detailed set of predictions by means of which the theory can be substantiated or rejected. Kiesler (1983) lists 18 references in support of his model. In a more recent review of the literature Orford (1986) comes to more cautious conclusions. The assertion that complementary responses are most common and anticomplementary responses least common is, according to Orford, well supported by empirical findings, as are more detailed predictions regarding the 'friendly' half of Kiesler's circle. However, two kinds of acomplementary responses (HD in response to HD and FD in response to HS) occur more frequently than predicted. Orford mentions relative status, setting, and level of development of a relationship as possible extra-theoretical factors that affect responses.

5.2 The "script"

5.2.1 An artificial world

In devising a conventional personality questionnaire the test constructor can soon get down to the business of writing items. In an interactive test, however, the items cannot as easily be written independently since they have to fit into the context of the unfolding interaction between test and testee. Therefore, before the items for an interactive test can be written, the test constructor has to decide on the "story line" or script that will link items together. The most obvious script for a test using

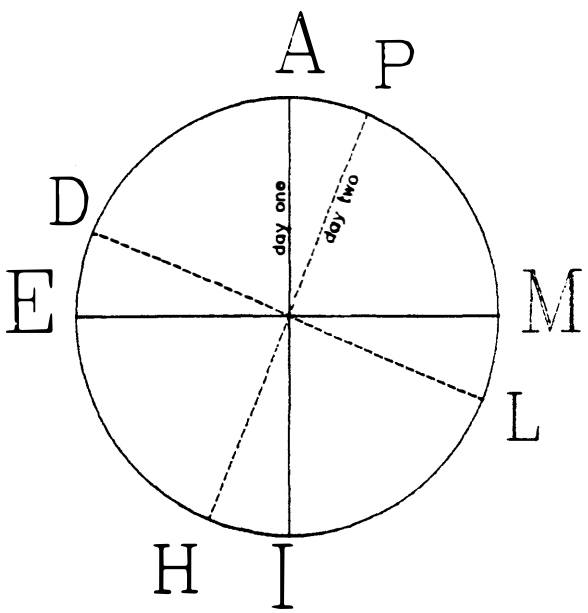
Kiesler's (1983,1985) circle would be that of a 'conversation' between computer and testee. However in Chapter 3 it was decided that this would be a dangerous course to follow since it could bring to the fore unpredictable feelings testees may have towards computers as "artificial beings". A less obvious script, playing itself out in an "artificial world" was therefore decided upon.

It was decided that the artificial world within which the RAP functions should be set in an educational institution since all people have had some experience with such institutions. An added advantage is that educational settings are often weakly or ambiguously structured, which facilitates the emergence of individual differences among testees (Amelang & Borkenau, 1986). The testee is put in the position of a new student at the Sullivan O'Leary College, a somewhat unusual place of learning where courses are offered in subjects such as ecology in addition to more usual 'school' subjects such as English. The intention is to make it possible for testees from various educational levels to identify with the college.

The story unfolds in the course of the testee's first four days at the college, during which he or she "meets" sixteen fictional characters. According to Kiesler's (1983) theory, people have "preferred styles" of interacting with other people which they tend to adopt in most circumstances. Each of the fictional people in the RAP has such a preferred style, corresponding to one of sixteen segments of the circle. During each of the four days at the college the testee interacts with four of the fictional people. An optimal sampling of the circle segments is

made on each day in that the four preferred styles of that day's fictional people are spaced evenly around the circle. On the first day, for instance, the fictional people come from segments A,E,I and M on the main axes of the circle; on day two the axes are rotated one position clockwise so that the fictional people have preferred styles at segments P,L,H and D (see Figure 5.3); on day three the axes are again rotated one position clockwise; and on day four once more.

Figure 5.3: Sampling of circle segments on days one and two



The fictional people include both blacks and whites. Both sexes and various age groups are also present. Each race and sex is equally represented on each day, and there is an equal number of people of each race and sex with preferred styles in each quadrant of the interpersonal circle. The fictional people are introduced to the testee by name and a sketch is provided of each. One such fictional person is depicted in Figure 5.4. Most of the sketches are fairly ambiguous with regard to the fictional

person's level of friendliness and dominance. The use of different sexes, races, names and sketches obviously can have a differential effect on testees' responses. This introduces a source of variance which is not directly related to what is being measured. The advantages are, however, likely to outweigh the disadvantages. An "artificial world" test is, as discussed, a simulation. The richer and more life-like the simulation, the better the results are likely to be. A sanitized single sex, single race, faceless, nameless artificial world could be constructed (e.g. a monastery where the monks wear masks and refer to each other as Brother A, Brother B, etc.), but this would be a poor reflection of reality.

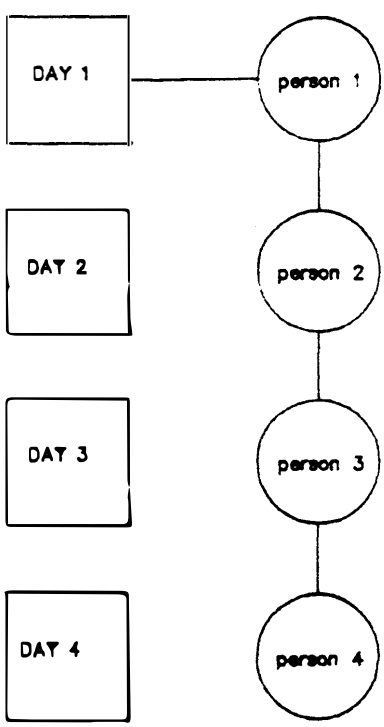
In order to have some control over the effects of contaminating variables such as sex, testees are required to make ratings of fictional people based on their names and sketches only. For purposes of comparison, a second rating is done at the end of an interaction with a fictional person.

Figure 5.4: A fictional person from the RAP

In an effort to sample some of the extratheoretical factors that affect interpersonal behaviour, the setting in which the interactions occur on each day was designed to 'push' testees towards one of the four poles of the interpersonal circle. On the first day the testee has to co-operate with fellow students in sharing scarce resources, thus possibly pushing him or her towards friendliness. On the second day he or she meets some of the lecturers; the status differential makes submissiveness an appropriate option. On the third day he or she is partially rejected by members of a club; this will evince hostile reactions in many people. And on day four he or she is appointed as co-ordinator of a project, thus possibly eliciting dominant behaviours from him or her.

The general outline of the artificial world is illustrated in Figure 5.5.

Figure 5.5: An outline of the artificial world



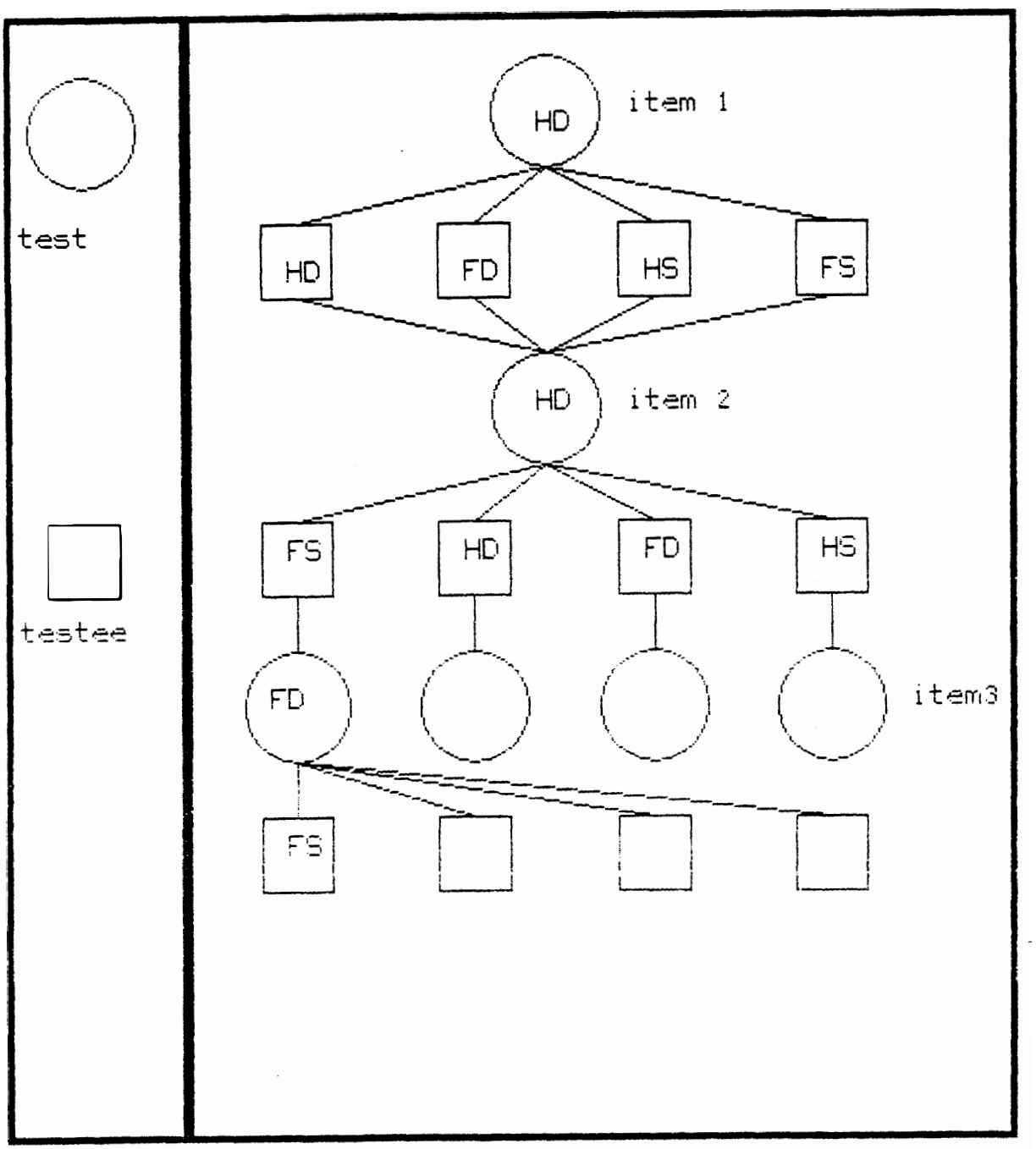
5.2.2 "Runs" of items

Each item in the RAP was written to represent one of the sixteen segments of the interpersonal circle, i.e. each item belongs to one of sixteen scales. Each item also occurs in one of sixteen "runs". It represents part of an exchange with one of the sixteen fictional people. The first two items in any run are standard for all testees. These two items represent the "preferred style" of the particular fictional person whose actions are described in the run of items. The testee must choose one of four ways of responding to the fictional person. The four response options represent four segments of the interpersonal circle - one from each quadrant. At the same time the options represent the four modes of responding identified by Kiesler (1983) - complementary, semimorphic, isomorphic, and anti-complementary. The fact that options have to be chosen from a list is limiting, but does prevent unwanted intrusions by the computer-as-being into the test world (as discussed in chapter 3).

The testee's response to the first item in a run has no effect on which item is presented next. After the second item, however, a branching sequence is set in motion, so that the item which is presented third represents the fictional person's response to the testee's behaviour. The same four options open to the testee are open to the fictional person and his or her potential responses can also be classified as representing four segments, four quadrants or the four modes of responding. Thus there are four

potential item 3's in each run, only one of which is 'realised' in any particular testing session. A hypothetical testee's path through a run (as far as item 3) is depicted in Figure 5.6.

Figure 5.6: A testee's path as far as item 3



In Figure 5.6 the fictional person is shown as responding in a complementary way in item 3 to the testee's response in item 2 (behaviour from quadrant FD constitutes a complementary response to behaviour from quadrant FS). The fictional person's action is not predetermined by the structure of the test, however, but is dependent on an algorithm that constructs a model of the testee and decides on the fictional person's reaction accordingly. Any algorithm could be used, based more or less closely on Kiesler's (1983) model. An algorithm dictating that friendly behaviour by the testee should be interpreted as hostility, although contrary to the theory, is perfectly possible. The script is deliberately flexible on this point so that different theories of how interaction occurs can be tried out simply by "plugging in" a different algorithm.

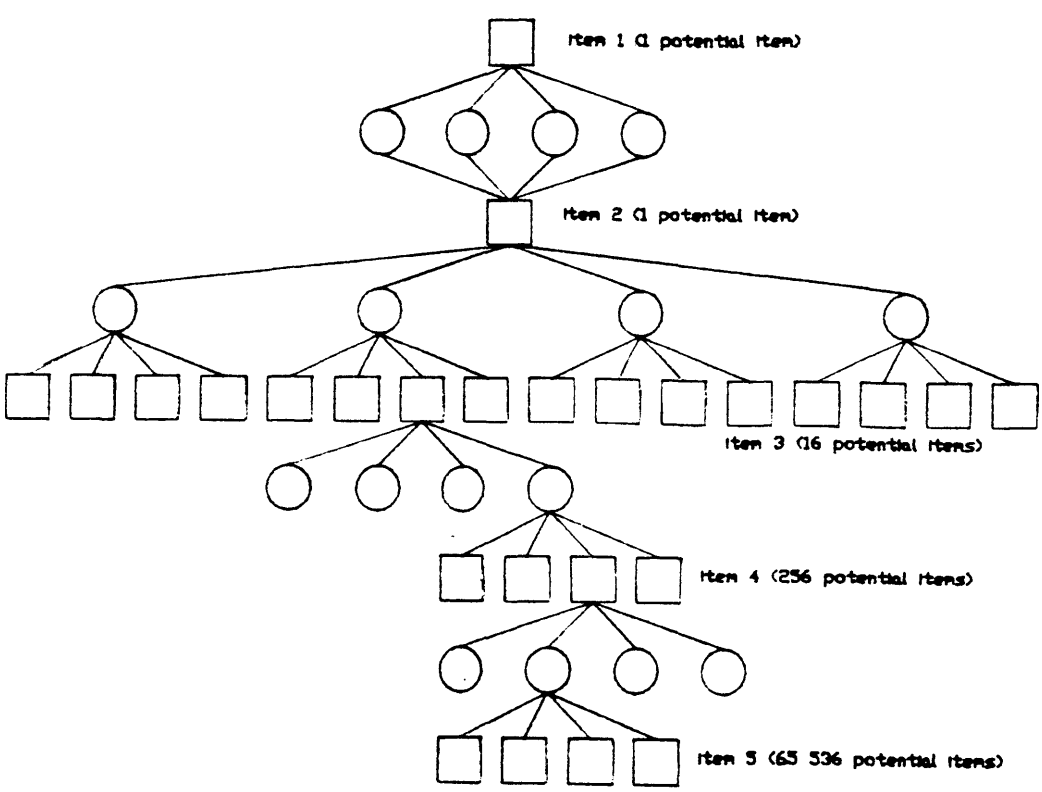
Items 4 and 5 in a run operate in the same way as item 3, various item selection algorithms again being a possibility. A very complicated algorithm for item 5 would construct a testee model based on all the information gleaned about the testee during the first four items, before deciding on which item 5 to present.

5.3 Construction of items

A detailed view of one run of items is given in Figure 5.7. As can be seen, only five items are actually presented in a run, but a very large number of "potential" items have to be available to make this possible (1+1+16+256+65 536=65 810). If

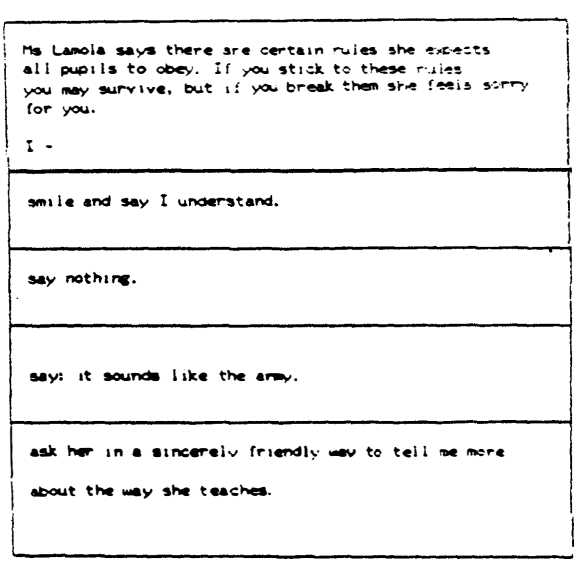
one considers that there are 16 runs, the number of items that have to be written becomes prohibitive. In order to cut down on this it was decided to 'collapse' potential items into each other. The sixteen potential item 3s, the 256 potential item 4s and the 65 536 potential item 5s were collapsed to four each. The four potential items at each level of a run represent four segments of the circle (one from each quadrant) as well as the four modes of responding, i.e. complementary, semimorphic etc. Items are written in such a way that any one of the four potential items at level three can plausibly lead to any of the four potential items at level four, and from there to any of the four potential items at level five. In addition, the response options for each of the potential items are identical.

Figure 5.7: One run of items



The number of items that had to be constructed was in this manner dramatically reduced, but at the cost of making the process of writing items considerably more difficult. Each item had to be carefully checked to see that it could fit into any one of the large number of paths that could lead into it. An example of an item is shown in Figure 5.8. The item shown is one of four possible item 3s in a run. The full run is given in Appendix A. The behaviour of the fictional person (Frank) described in this particular item comes from the HS quadrant of the interpersonal circle. This item, rather than equivalent items representing quadrants HD, FD or FS, was selected by the algorithm as an appropriate response to the testee's behaviour in item 2. The testee's response options in this item represent four segments and four quadrants of the interpersonal circle. Response 1 is from the FS quadrant, response 2 is HD, response 3 is FD, and response 4 is HS. The responses can also be classified as complementary (response 2), anti-complementary (response 1), semimorphic (response 3) and isomorphic (response 4).

Figure 5.8: An item similar to items in the RAP



Item construction was done by means of the deductive (or "armchair") method. A review by Burisch (1986) suggests that this method may have the edge (as far as discriminant validity and economy is concerned) over other methods such as factor analysis. The actual writing of items was greatly facilitated by an overt or behavioural version of the circle developed by Kiesler (1985). The preferred style of a fictional person and the possible deviations from that style were first carefully studied and a mental picture of the fictional person built up. The fictional person's position in the script was next considered and a list drawn up of events that could possibly occur with the person at that point in the script. Next, a general outline was written of how an interaction with this person could unfold. Finally items were written with reference to this outline. An effort was made to include as many as possible of Kiesler's list of behaviours for a particular segment, but the sampling is not claimed to be exhaustive. Only behaviours rated by Kiesler as of moderate intensity were included. Very intense (or extreme) behaviours is indicative of pathology, whereas the RAP is intended primarily as a test of normal functioning. Having more than one level of intensity would moreover have complicated the branching structure of the test unduly.

On the whole the resultant items appear not to be overly subtle. They are couched in overt behavioural form, with little reference to covert feelings. There are few references to out-of-focus behaviours such as body language. The overt

confirms a finding by Angleitner, John and Lohr (1986) that items measuring interpersonal traits "tend to be conceptualized in terms of overt behavioral reactions" (p. 76). The relative obviousness of the items reflects the conclusion (reached in Chapter 3) that there is little to be gained from making items extremely subtle.

The items were written and revised by the author after trying them out on colleagues. Comment was obtained from independent judges in the Assessment and Counselling division of the NIPR. Many of these judges balked at the large number of potential items that had to be assessed, however. They found it a very tedious task to trace each of the possible paths that could lead to an item in order to get an idea of its context. It would seem that the initial writing of items for an interactive test is necessarily a more individual activity than for a conventional test. Items in an interactive test cannot be assessed out of context and a panel of judges cannot be expected laboriously to trace the possible contexts of each item. It is possible that the use of "objective" judges can be dispensed with entirely. Burisch (1986) argues that the quality of items depends mostly on the accuracy of the author's intuition and can be only marginally improved by the use of a panel of judges.

5.4 Scoring

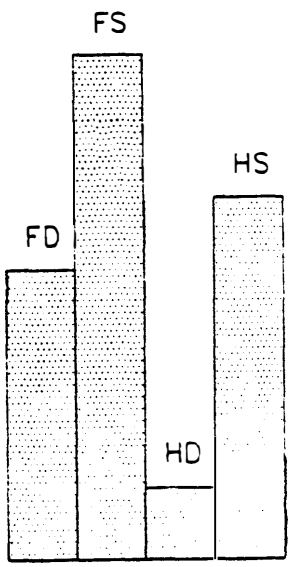
When constructing a conventional personality test the question as to how it should be scored rarely arises. Although unusual scoring methods have been advocated (e.g. by Jannarone & Roberts, 1984), the standard practice is simply to add up item unit scores to obtain scale scores. In constructing a new kind of test such as the RAP, however, the question of scoring is crucial. If new kinds of information is to be gleaned from the test, it must be clear in advance how this can be done.

Seven forms of scoring the RAP have been developed. These are described below as an illustration of the extent to which the information obtainable from an interactive test goes beyond that obtainable from conventional tests.

The first two scoring methods are conventional. The first is to simply add up the number of times a testee chooses an option from each of the four quadrants of the interpersonal circle <u>in the first item of each run.</u> This is the equivalent of turning the RAP into a sixteen-item conventional test, with each item being an independent unit. This kind of scoring is possible because the degree to which branching occurs was limited in accordance with the recommendation in Chapter 4. If the interaction had been allowed to meander on without a break from beginning to end, instead of being broken up into 16 runs, this kind of scoring would not have been possible. Although not likely to be very reliable, this score could serve as a useful benchmark.

The second kind of scoring is to add up equally weighted responses to all items and thus to get summary scores of the number of responses in each quadrant and segment of the interpersonal circle. A sample score distribution using this kind of scoring is depicted in Figure 5.9. At first this kind of scoring seems simple-minded in the extreme: An adding-up procedure assumes that items are functionally independent, which is explicitly not the case here. Worse still, testees are not presented with the same items, making their scores not strictly comparable. The rationalization for nevertheless using this procedure was mentioned in Chapter 4, namely that an interactive test could be seen as a simulation of real life. To the extent that items are not independent in the test, situations are also not independent in real life; to the extent that testees are presented with different items, they are also presented with different situations in real life. A person may for instance, through his behaviour in the test, cause a large number of hostile items to be presented to him, which is likely to elicit even further hostility. However, if the test simulates life accurately this kind of vicious circle is exactly what occurs in the person's everyday life.

Figure 5.9: Sample scores using scoring method 2.



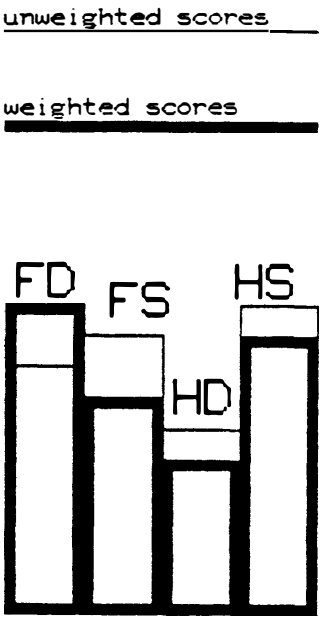
The success of the second scoring method clearly depends on the extent to which the RAP simulates real life. The testee needs to experience the entire situation as realistic. The branching algorithm in particular has to capture the way in which real life events lead on to one another.

Scoring methods 3 to 7 are less conventional than the first two. These take account of the context within which a response occurs. Thus the unit of measurement becomes the transactions between the test and the testee, as discussed in the previous chapter. Method 3 is to weight responses in terms of how predictable they are according to Kiesler's (1983) theory. It will be remembered that Kiesler identifies four types of responses. The most likely are complementary responses where the person accepts both the control and the affiliation bids of the other person. If for instance the first person acts friendly-dominant (FD) way and the second person acts in a friendly-submissive (FS) way, the transaction is complementary. The least likely response is an anti-complementary response, where both the affiliation and the control bids are rejected. Semimorphic and isomorphic responses are intermediate in that either the control or the affiliation bids are rejected, but not both.

If a person acts in a way thought by Kiesler (1983) to be unlikely (e.g. he produces an anti-complementary response), this could be because Kiesler's theory is wrong or not applicable to the person, but it could also be because the likely response was strongly contrary to the person's preferred style. Somebody who

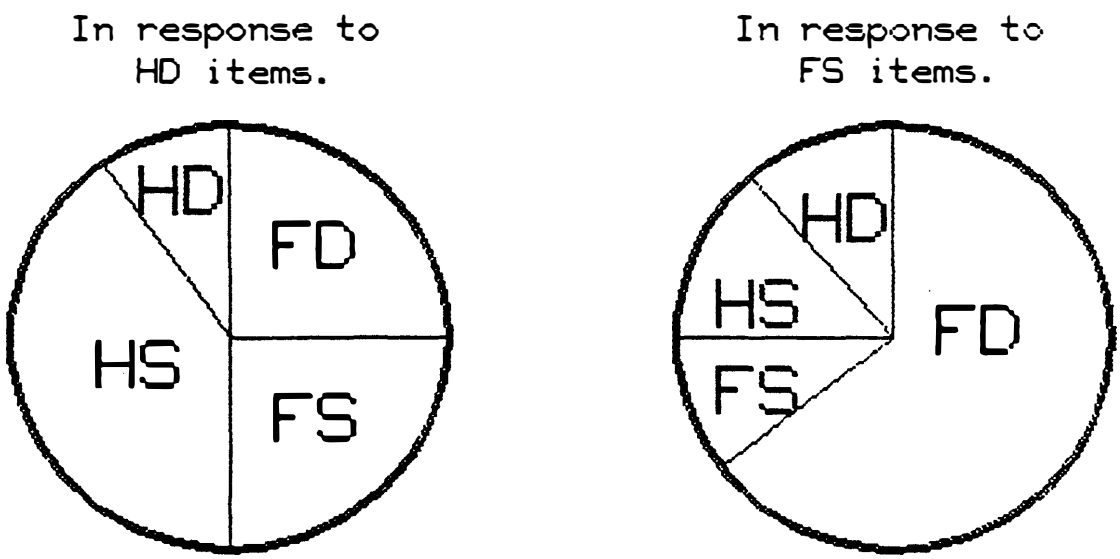
finds it very hard to act in a hostile or a dominant manner may, when faced with hostile-submissive behaviour, fail to produce the expected complementary response of hostile-dominance, but instead act in an anti-complementary way (hostile-submissive). Thus when a person responds in a complementary way he or she is merely doing what is expected of people in social situations and his or her behaviour may tell us little about his or her preferred style. In scoring method 3 this kind of response is therefore weighted only 1. A person who responds in an anti-complementary manner reveals more about him- or herself, and this kind of response is accordingly weighted 3. Semimorphic and isomorphic responses are weighted 2. The difference between a testee's weighted and unweighted scores are shown in Figure 5.10. The bigger the difference, the less suitable the branching algorithm was for a testee. The success of the third scoring method thus does not depend as much on a good branching algorithm, but still depends on Kiesler's theory being accurate and on the items accurately reflecting the theory.

Figure 5.10: Sample weighted vs unweighted scores.



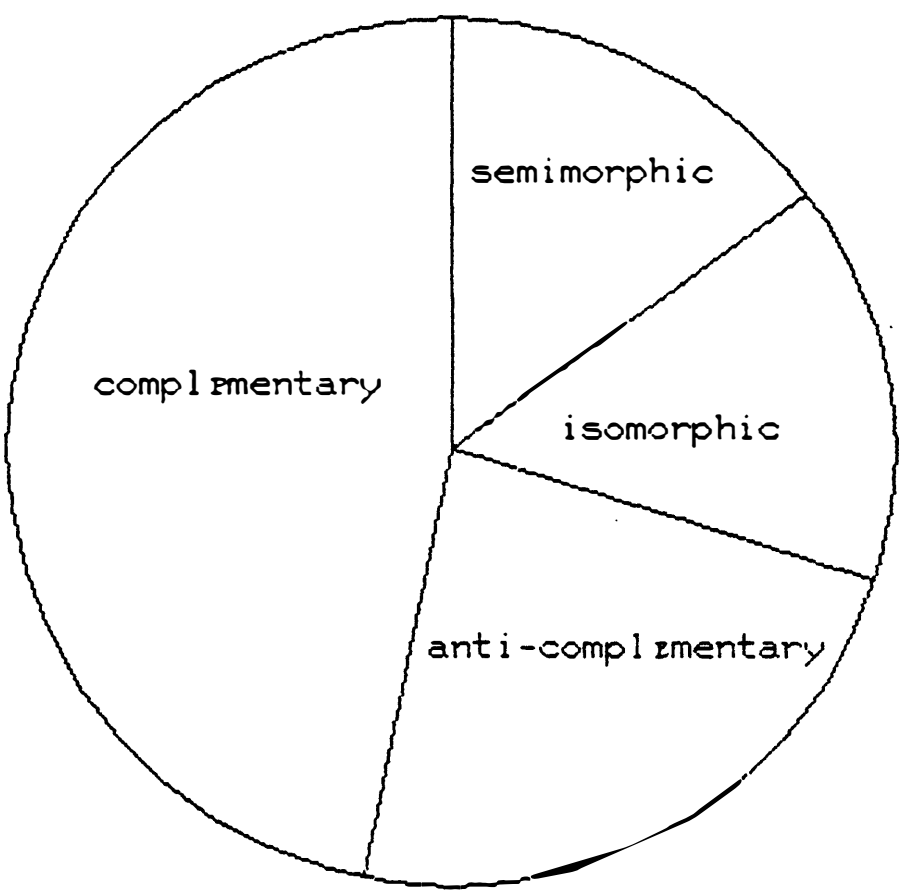
The fourth scoring method is simply to give scores as conditional statements (e.g. a testee responded with hostility to friendliness so many times, with hostility to hostility so many times and so on). This method reflects the need, identified in Chapter 4, to be able to make conditional statements about people-in-situations. This kind of scoring makes more fine-grained prediction possible than conventional test scores. It may for instance be possible to predict whether a person is likely to get along with a dictatorial boss or not. An example of such conditional scores is shown in Figure 5.11.

Figure 5.11: Sample conditional scores.



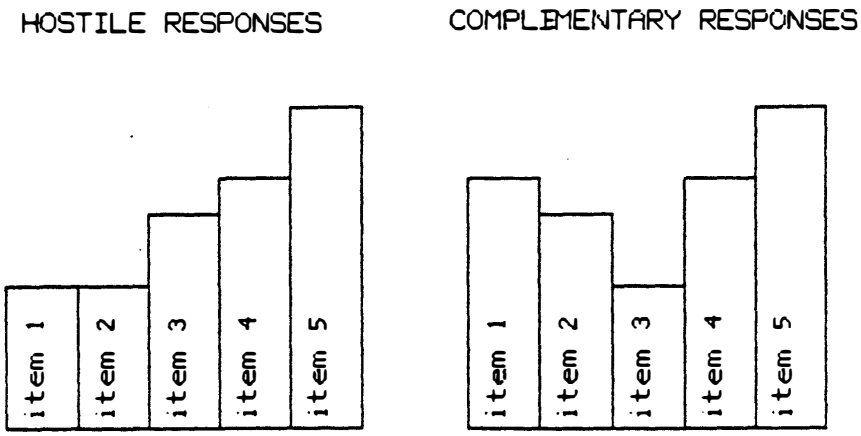
The fifth scoring method is to use the context in which responses occur (i.e. whether they are complementary, anti-complementary, semimorphic and isomorphic responses) as indices of interest in their own right, not just as moderators for other scores. A person with a high number of complementary responses is, for instance, likely to be capable of conventional interpersonal behaviour in real life, whereas somebody with many anti-complementary responses is likely to be a "maverick" or lack social competence. It is also possible that he or she may not have understood the items. Sample scores of this kind are shown in Figure 5.12.

Figure 5.12: Sample scores using method 5.



The sixth scoring method is intended to show how responses changed over time in the course of the five items that constitute each run. One could for instance imagine a person who tends to start off a relationship quite submissively (i.e. responds submissively to items 1 and 2 in each run), but gradually builds up to become more dominant (i.e. tends to respond dominantly to item 5). The number of complementary responses may also vary across time, being most frequent at the beginning, when conventional behaviour is a safe option, and at the end, when the relationship has settled down. Orford's (1986) observation that Kiesler's (1983) predictions may function differently depending on the time in a relationship may thus be tested. A run of five items unfortunately does not give much opportunity for a relationship to be established between a testee and a fictional person. The possibility mentioned in the previous chapter that valid variance is likely to be lost in direct proportion to the degree that branching is limited thus seems correct. Lengthening one or more of the runs in the test must therefore be considered for the future. Examples of the sixth type of score are shown in Figure 5.13.

Figure 5.13: Sample responses over time.



The last scoring method moves furthest away from Kiesler's (1983) system. It is based on Tracey's (1985) method of determining who is most dominant in an interaction. Tracey proposes statistical predictability as a criterion for dominance: "In a dyad, if B's future behavior is more predictable from A's past behavior than conversely, then A is said to be dominant" (p. 119). Although Tracey uses topic changing, the method is not in principle limited to any particular kind of behaviour. In the RAP, for instance, one could use any of Kiesler's classifications of behaviour (e.g. dominant or friendly-submissive or semimorphic behaviour). Should one decide for instance to use friendliness as the behaviour of interest, the relative predictability of the testee (T) and the fictional person (FP) would be computed as a function of a) the frequency of FP friendliness following T friendliness, b) the frequency of FP hostility following T hostility, c) the frequency of FP hostility following T friendliness, and d) the frequency of FP friendliness following T hostility. The kind of dominance identified by this procedure is not the same as conventional trait-dominance. That it exists in real life seems likely, though. It is not hard to imagine the kind of person who on the surface acts in a very submissive manner, but who nevertheless always manages to make people act in predictable ways by his or her very submissiveness.

Seven different ways of scoring responses to RAP-items have been proposed. More may be possible. The methods suggested range from conventional, to quite unusual. The scoring methods are suggestive of the richness of information that can be extracted

from an interactive test - information that goes well beyond the simple trait-scores yielded by conventional tests.

5.5 Conclusion

In this chapter the construction of the Relational Adaptation Profile (RAP) was described. The intention was to illustrate in a concrete manner the issues discussed in the previous chapters. It is hoped that it could thus be demonstrated that the use of computer technology holds some promise for innovation in personality assessment. The validation of the RAP will be described in a forthcoming report.

6. FINAL CONCLUSION

In this report the prospects for innovation in personality assessment were discussed. The discussion was presented more or less in the form of an argument leading from general premises to quite specific conclusions. These conclusion were then implemented in the construction of a new interactive test. In tracing out the line of the argument it was inevitable that along the way promising possibilities would be overlooked or brushed aside in the interests of coherence. It is hoped that this is offset by the degree of in-depth attention that could be brought to bear on those possibilities that <u>were</u> considered.

The bare bones of the argument around which this report was structured are the following: Personality assessment has reached a point of stagnation. Three possible routes out of the present impasse involve personality theory, psychometrics and technology. Neither personality theory nor psychometrics hold much promise - the former because it does not map directly onto assessment procedures, the latter because it tends to get bogged down in detail at the expense of the bigger picture. Technology is more promising, particularly with the advent of computing-for-the-masses. Previous analyses of the implications of computer technology for personality assessment are not very convincing. A more useful point of departure for such an analysis is to view technology as playing a mediational role between tester and testee. In terms of this role, personality tests can be classified according to their degree of interactivity (games vs dreams), the manner in which delegation occurs (beings vs worlds) and their subtlety (questions vs koans). Computer technology can best be used for the improvement of personality assessment if tests are constructed as games, worlds and questions. Of the three dimensions of the assessment relationship, the degree of interactivity is most strongly affected by the use of computers. The unique advantage that computers have over other forms of assessment technology is their interactive capability. An interactive assessment situation is characterized by the reciprocal causality between the testee's and the test's actions. Conventional psychometrics and personality theory cannot deal adequately with reciprocal causality. Consequently, to make full use of computer technology changes would have to occur in both the personality theory and

the psychometrics usually associated with personality assessment. In the former case a move towards more interactional theories, in particular interpersonal theory, is required; in the latter case either pragmatic solutions or more appropriate statistics such as process statistics will have to be developed and applied.

The interactive test presented as the culmination of the report is, it is hoped, suggestive of the prospects for innovation in personality assessment offered by computer technology. The dream of the computer as interviewing psychologist is not yet within our reach, and may never be, but significant advances over currently used instruments are now distinctly possible.

REFERENCES

Allen, R.B. (1982). Cognitive factors in human interaction with computers. In A. Badre & B. Schneiderman (Eds.) Directions in human/computer interaction. Norwood: Ablex Publishing Corporation

Amelang, M. & Borkenau, P. (1986). The trait concept: Current theoretical considerations, empirical facts, and implications for personality inventory construction. In A. Angleitner & J.S. Wiggens (Eds.) Personality Assessment via Questionnaires pp. 7 - 34. Berlin: Springer-Verlag.

Anastasi, A. (1982). Psychological Testing. New York: Macmillan.

Anchin, J.C. (1982). Sequence, pattern and style: Integration and treatment implications of some interpersonal concepts. In J.C. Anchin & D.J. Kiesler (Eds.) Handbook of interpersonal psychotherapy pp. 95 - 131. New York: Pergamon Press.

Anchin, J.C. & Kiesler, D.J. (Eds.) (1982). Handbook of interpersonal psychotherapy. New York: Pergamon Press.

Andrich, D. (1985). A latent-trait model for items with response dependencies: implications for test construction and analysis. In S.E. Embretson (Ed.) Test Design. Orlando: Academic Press.

Angleitner, A., John, O.P. & Lohr, F.J. (1986). It's what you ask and how you ask it: an itemmetric analysis of personality questionnaires. In A. Angleitner & J.S. Wiggens (Eds.) Personality Assessment via Questionnaires pp. 61 - 108. Berlin: Springer-Verlag.

Angleitner, A. & Wiggens, J.S. (Eds.) (1986). Personality Assessment via Questionnaires. Berlin: Springer-Verlag.

Angyal, A. (1981). Personality as a hierarchy of systems. In F.E. Emery (Ed.) Systems Thinking (Vol. 2) pp. 125 - 137. Middlesex: Penguin Books.

Ayton, P. & Wright, G. (1985). The evidence for interactionism in psychology: a reply to Furnham and Jaspers. Personality and Individual Differences, 6, 509 - 512.

Badre, A. & Schneiderman, C. (Eds.) (1982). Directions in human/computer interaction. Norwood: Ablex Publishing Corporation.

Barker R.G. (1968). Ecological Psychology. Stanford: Stanford University Press.

Bartram, D. & Bayliss, R. (1984). Automated testing: Past, present and future. Journal of Occupational Psychology, 57, 221 - 237.

Bateman, J.A. (1985). The role of language in maintenance of intersubjectivity: a computational investigation. In G.N. Gilbert & C. Heath (Eds.) Social Action and Artificial Intelligence pp. 40 - 81. Aldershot: Gower Publishing Company.

Beaumont, J.G. (1981). Microcomputer-aided assessment using standard psychometric procedures. Behavior Research Methods & Instrumentation, 13, 430 - 433.

Bobrow, D.G. & Hayes, P.J. (1985). Artificial Intelligence - where are we. Artificial Intelligence, 25(3), 375 - 415.

Boguraev, B. (1985). User modelling in cooperative natural language front ends. In G.N. Gilbert & C. Heath (Eds.) Social Action and Artificial Intelligence pp. 124 - 143. Aldershot: Gower Publishing Company.

Burisch, M. (1986). Methods of personality inventory development - a comparitive analysis. In A. Angleitner & J.S. Wiggens (Eds.) Personality Assessment via Questionnaires pp. 109 - 120. Berlin: Springer-Verlag.

Burkhart, B.R., Christian, W.L. & Gynther, M.D. (1978). Item subtlety and faking on the MMPI: A paradoxical relationship. Journal of Personality Assessment, 42(1), 76 - 80.

Buss, D.M. & Craik, K.H. (1983). Act prediction and the conceptual analysis of personality scales: Indices of act density, bipolarity, and extensity. Journal of Personality and Social Psychology, 45, 1081 - 1095.

Campbell, D., Reder, P., Draper, R. & Pollard, D. (1982). Working with the Milan method: twenty questions. Occasional papers on family therapy No. 1. London: Institute of Family Therapy.

Capra, F. (1976). The Tao of Physics. Bungay: Fontana.

Cartwright, D.S. (1975). Trait and other sources of variance in the S-R inventory of anxiousness. Journal of Personality and Social Psychology, 32, 408 - 414.

Claeys, W., De Boeck, P., Van den Bosch, W., Biesmans, R. & Bohrer, A. (1985). A comparison of one free-format and two fixed-format self-report personality assessment methods. Journal of Personality and Social Psychology, 49, 1028 - 1039.

Conte, H.R. & Plutchik, R. (1981). A circumplex model for interpersonal personality traits. Journal of Personality and Social Psychology, 40, 701 - 711.

Cook, J.T. (1983). The use of selected communication media in structured empathy training. Unpublished MA thesis, University of South Africa, Pretoria.

Cook, J.T. (1984). Empathy from the perspective of systems theory. Paper read at 2nd PASA congress. Pretoria: HSRC.

Coulter, J. (1985). On comprehension and 'mental representation'. In G.N. Gilbert & C. Heath (Eds.) Social Action and Artificial Intelligence pp. 8 - 23. Aldershot: Gower Publishing Company.

Coulter, M.A. (1973). Item Response Evaluation. CSIR Special Report PERS 186. Johannesburg: National Institute for Personnel Research.

Cronbach, L.J. (1984). Essentials of Psychological Testing. (Fourth Edition). New York: Harper & Row.

Davis, D. & Perkowitz, W.T. (1979). Consequences of responsiveness in dyadic interaction: effects of content-related responses on intepersonal attraction. Journal of Personality and Social Psychology, 37, 534 - 550.

Dreyfus, H.L. (1979). What computers can't do. New York: Harper Row.

Dreyfus, H.L. & Dreyfus, S.E. (1986). Mind over machine. The power of human intuition and expertise in the era of the computer. New York: The Free Press.

Dubinsky, S., Gamble, D.J. & Rogers, M.L. (1985). A literature review of subtle-obvious items on the MMPI. Journal of Personality Assessment, 49 (1), 62 - 68.

Duncan, S., Kanki, B.G., Mokros, H. & Fiske, D.W. (1984). Pseudounilaterality, simple-rate variables, and other ills to which interaction research is hier. Journal of Personality and Social Psychology, 46, 1335 - 1348.

Emery, F.E. (Ed.) (1981) . Systems Thinking (Vol 1). Middlesex: Penguin Books.

Emmons, R.A., Diener, E. & Larsen, R.J. (1985). Choice of situations and congruence models of interactionism. Personality and Individual Differences, 6(6), 693 - 702.

Endler, N.S. & Hunt, J.M. (1966). Sources of behavioral variance as measured by the S-R inventory of anxiousness. Psychological Bulletin, 65, 336 - 346.

Endler, N.S. & Hunt, J.M. (1969). Generalizability of contributions from sources of variance in the S-R inventories of anxiousness. Psychological Bulletin, 37, 1- 24.

Endler, N.S., Hunt, J.M. & Rosenstein, A.J. (1962). An S-R inventory of anxiousness. Psychological Monographs, 76 (17), 1 - 33.

Epstein, S. (1979). The stability of behavior: I. On predicting most of the people much of the time. Journal of Personality and Social Psychology, 37, 1097 - 1126.

Erdman, H.P., Greist, J.H., Klein, M.H., Jefferson, J.W. & Getto,C. (1981). The computer psychiatrist: How far have we come? Where are we heading? How far dare we go? Behavior Research Methods & Instrumentation, 13, 393 - 398.

Erdman, H., Klein, M.H. & Greist, J.H. (1983). The reliability of a computer interview for drug use/abuse information. Behavior Research Methods & Instrumentation, 15, 66 - 68.

Feigenbaum, E.A. & McCorduck, P. (1984). The Fifth Generation: Artificial Intelligence and Japan's computer challenge to the world. London: Pan Books.

Forester, T. (Ed.) (1984). The Microelectronics Revolution. Oxford: Basil Blackwell.

Friedlander, M.L. & Phillips, S.D. (1984). Stochastic Process Analysis of interactive discourse in early counselling interviews. Journal of Counseling Psychology, 31(2), 139 - 148.

Furnham ,A & Jaspers, J. (1985). More confusion. Personality and Individual Differences, 6(4), 513 - 514.

Fyans, L.J. (Ed.) (1980). Achievement Motivation: recent trends in theory and research. New York: Plenum Press.

Gilbert, G.N. & Heath, C. (Eds.) (1985). Social Action and Artificial Intelligence. Aldershot: Gower Publishing Company.

Good, D. (1985). Sociology and AI: The lesson from social psychology. In G.N. Gilbert & C. Heath (Eds.) Social Action and Artificial Intelligence pp. 82 - 103. Aldershot: Gower Publishing Company.

Gorta, A. (1985). Choosing situations for a purpose. European Journal of Social Psychology, 15(1), 17 - 35.

Gough, H.G. (1956). California Psychological Inventory. Palo Alto: Consulting Psychologists Press.

Green, T.R.G., Payne, S.J. & Van der Veer, G.C. (Eds.) (1983). The Psychology of computer use. London: Academic Press.

Greenberg, L.S. (1986). Change process research. Journal of Consulting and Clinical Psychology, 54(1), 4 - 9.

Grygier, T.G. (1956). Dynamic Personality Inventory. London: National Foundation for Educational Research in England and Wales.

Hasset, J. (1984). Computers in the classroom. Psychology Today, 18(9), 22 - 28.

Hathaway, S.R. & McKinley, J.C. (1951). Minnesota Multiphasic Personality Inventory Manual. New York: The Psychological Corporation.

Hess, A.K. & Neville, D. (1977). Testwiseness: Some evidence for the effect of personality testing on subsequent test results. Journal of Personality Assessment, 41(2), 170 - 177.

Hiebsch, H., Brandstatter,H. & Kelley, H.H. (Eds.) (1982). Social Psychology. Amsterdam: North-Holland Publishing Company.

Hofstadter, D.R. (1979). Godel, Escher, Bach: An eternal golden Braid. New York: Vintage Books.

Holden, R.R. & Jackson, D.N. (1985). Disguise and the structured self-report assessment of psychopathology: I. An analogue investigation. Journal of Consulting and Clinical Psychology, 53(2), 211 - 222.

Hunt, D.E. (1980). From single-variable to persons-in-relation. In L.J. Fyans (Ed.), Achievement Motivation: Recent trends in theory and research. New York: Plenum Press.

Hunt, E. & Pellegrino, J. (1984). Using interactive computing to expand intelligence testing. A critique and prospectus. Technical Report 84-2. Santa Barbara: University of California.

Jackson, D.N. & Helmes, E. (1979). Personality structure and the circumplex. Journal of Personality and Social Psychology, 37, 2278 - 2285.

Jannarone, R.J. & Roberts, J.S. (1984). Reflecting interactions among personality items: Meehl's paradox revisited. Journal of Personality and Social Psychology, 47(3), 621 - 628.

Johnson, J.A. (1981). The "self-disclosure" and "self-presentation" views of item response dynamics and personality scale validity. Journal of Personality and Social Psychology, 40, 761 - 769.

Johnson, J.H., Harris, W., Thompson, C. Marcus, S., Block, D., Novak, K., Yingst, N., Allert, A., Hobin, G., Byrnes, E., Strauch, D., Feldman, C., Niedner, D. & Fink, A. (1981). New uses for the on-line computer medium in mental health care. Behavior Research Methods & Instrumentation, 13, 243 - 250.

Johnson, J.H. & Johnson, K.N. (1981). Psychological considerations related to the development of computerized testing stations. Behavior Research Methods & Instrumentation, 13, 421 - 424.

Jordaan, W.J. & Jordaan, J.J. (1980). Metateorie: 'n Sesde 'krag' in die sielkunde. South African Journal of Psychology, 10(1/2), 28 - 41.

Keeny, B.P. (1979). Ecosystemic epistemology: an alternative paradigm for diagnosis. Family Process, 18(2), 117 - 129.

Keeny, B.P. & Morris, J. (1985). Implications of cybernetic epistemology for clinical research: A reply to Howard. Journal of Counseling and Development, 63(9), 548 - 550.

Kelly, G.A. (1958). The theory and technique of assessment. Annual Review of Psychology, 9, 323 - 352.

Kenny, D.A. & LaVoie, L. (1985). Separating individual and group effects. Journal of Personality and Social Psychology, 48(2), 339 - 348.

Kiesler, D.J. (1982). Interpersonal Theory for Personality and Psychotherapy. In J.C. Anchin & D.J. Kiesler (Eds.) Handbook of interpersonal psychotherapy. New York: Pergamon Press.

Kiesler, D.J. (1983). The 1982 Interpersonal Circle: A taxonomy for complimentarity in human transactions. Psychological Review, 90(3), 185 - 214.

Kiesler, D.J. (1985). The 1982 Interpersonal Circle: Acts version. Richmond: Virginia Commonwealth University.

Kingston, N.M. & Dorans, N.J. (1982). The effect of the position of an item within a test on item responding behavior: An analysis based on item response theory. GRE Board Professional Report 79-12bP. Princeton: Educational Testing Service.

Kingston, N.M. & Dorans, N.J. (1984). Item location effects and their implications for IRT equating and adaptive testing. Applied Psychological Measurement, 8(2), 147 - 154.

Kleinmuntz, B. (1967). Personality Measurement. Homewood: The Dorsey Press.

Kleinmuntz, B. & McLean, R. (1968). Diagnostic interviewing by digital computer. Behavioral Science, 13, 75 - 80.

Kopp, S. (1974). If you meet the Buddha on the road, kill him! London: Sheldon.

Koson, D., Kitchen, C., Kochen, M. & Stodolosky, D. (1970). Psychological testing by computer: Effect on response bias. Educational and Psychological Measurement, 30, 803 - 810.

Krauss, R.M., Apple, W., Morency, N., Wenzel,C. & Winton, W. (1981). Verbal, vocal, and visible factors in judgements of another's affect. Journal of Personality and Social Psychology, 40, 312 - 320.

Kunce, J. & Kunce, H. (1982). Robots, simulation theory, and human interpersonal interactions. In H. Hiebsch, H. Brandstatter, & H.H. Kelly (Eds.) Social Psychology pp. 215 - 221. Amsterdam: North-Holland Publishing Company.

Kuncel, R.B. (1973). Response processes and relative location of subject and item. Educational and Psychological Measurement, 33, 545 - 563.

Lanyon, R.I. (1984). Personality Assessment. Annual Review of Psychology, 35, 667 - 701.

Lanyon, R.I. & Goodstein, L.D. (1982). Personality Assessment. New York: John Wiley & Sons.

Leary, T. (1957). Interpersonal Diagnosis of Personality. New York: The Ronald Press Company.

Lewin, K. (1935). A dynamic theory of personality. New York: McGraw-Hill.

Liebert, R.M. & Spiegler, M.D. (1982). Personality. Strategies and issues. Homewood: The Dorsey Press.

Lyons, J.P. & Brown, J. (1981). Reduction in clinical assessment time using computer algorithms. Behavior Research Methods & Instrumentation, 13, 407 - 412.

Maass, S. (1983). Why systems transparency? In Green, T.R.G., Payne, S.J. & Van der Veer, G.C. (Eds.) The Psychology of Computer Use pp. 19 - 28. London: Academic Press.

Magara, P.A., Ashbrook, R.M., Lesowitz, T.E. & Johnson, M.H. (1985). The consistency of hysteric traits across three situational contexts. Personality and Individual Differences, 6, 425 - 428.

McGrath, R.E., O'Malley, W.B. & Dura, J.R. (1986). Alternative scoring system of repeated items on the MMPI: Caveat Emptor. Journal of Personality Assessment, 50(2), 182 - 185.

McTear, M. (1985). Breakdown and repair in naturally occuring conversation and human-computer dialogue. In G.N. Gilbert & C. Heath (Eds.) Social Action and Artificial Intelligence pp. 104 - 123. Aldershot: Gower Publishing Company.

Miller, J.G. (1978). Living Systems. New York: McGraw Hill.

Michie, D. (Ed.) (1982). Introductory readings in expert systems. New York: Gordon and Breach Science Publishers.

Minsky, M. (1974). A framework for representing knowledge. Cambridge (Mass.): Artificial Intelligence Laboratory.

Mischel, W. (1968). Personality and assessment. New York: Wiley.

Morf, M., Alexander, P. & Fuerth, T. (1981). Fully automated psychiatric diagnosis: Some new possibilities. Behavior Research Methods & Instrumentation, 13, 413 - 416.

Murray, H.A. (1938, 1962). Explorations in personality. New York: Oxford University Press.

Natale, M. (1975). Convergence of mean vocal intensity in dyadic communication as a function of social desirability. Journal of Personality and Social Psychology, 32, 790 - 804.

Orford, J. (1986). The rules of interpersonal complimentarity: Does hostility beget hostility and dominance, submission? Psychological Review, 93(3), 365 - 377.

Osberg, T.M. (1985). Order effects in the administration of personality measures: the case of the Self-Consciousness Scale. Journal of Personality Assessment, 49(5), 536 - 540.

Oettinger, A.G. (1969). Run, computer, run. The mythology of educational innovation. New York: Collier Books.

Perkins, J.E. & Goldberg, L.R. (1964). Contextual effects on the MMPI. Journal of Consulting Psychology, 28, 133 -140.

Pinsky, L. (1983). What kind of "dialogue" is it when working with a computer? In Green, T.R.G., Payne, S.J. & Van der Veer, G.C. (Eds.). The Psychology of Computer Use pp. 29 - 40. London: Academic Press.

Prenis, J. (1981). The language of computers. London: W.H. Allen.

Regan, T. & Singer, P. (Eds.) (1976). Animal rights and human obligations. Englewood Cliffs: Prentice Hall.

Richards, J.S., Fine, P.R., Wilson, T.C. & Rogers, J.T. (1983). A voice-operated method for administering the MMPI. Journal of Personality Assessment, 47(2), 167 -170.

Rosenthal, D.A. & Lines, R. (1978). Handwriting as a correlate of extraversion. Journal of Personality Assessment, 42(1), 45 - 48.

Royce, J.R. (1979). Toward a viable theory of individual differences. Journal of Personality and Social Psychology, 37, 1927 - 1931.

Russel, R.L. & Trull, T.J. (1986). Sequential analyses of language variables in psychotherapy process research. Journal of Consulting and Clinical Psychology, 54(1), 16 - 21.

Sackman, H. (1967). Computers, system science, and evolving society. The challenge of man-machine digital systems. New York: John Wiley.

Schachtel, E.G. (1966). Experiential foundations of Rorschach's test. London: Tavistock Publications.

Schubert, S.P. & Fiske, D.W. (1973). Increase of item response consistency by prior item response. Educational and Psychological Measurement, 33, 113 - 121.

Schutz, W. (1977). FIRO-B. Palo Alto: Consulting Psychologists Press.

Searight, H.R. & Openlander, P. (1984). Systemic Therapy: A new brief intervention model. The Personnel and Guidance Journal, March, 387 - 391.

Selmi, P.M., Klein, M.H., Greist, J.H., Johnson, J.H. & Harris, W.G. (1982). An investigation of computer-assisted cognitive-behavior therapy in the treatment of depression. Behavior Research Methods & Instrumentation, 14, 181 - 185.

Shelley, M. (1978). Frankenstein. New York: Dale Books.

Silverman, I. (1977). The human subject in the Psychological laboratory. New York: Pergamon Press.

Simon, H.A. (1984). What computers mean for man and society. In Forester, T. (Ed.) The Microelectronics Revolution pp. 419 - 433. Oxford: Basil Blackwell.

Solomon, E. & Kopelman, R.E. (1984). Questionnaire format and scale reliability: an examination of three modes of item presentation. Psychological Reports, 54, 447 - 452.

Space, L.G. (1981). The computer as psychometrician. Behavior Research Methods & Instrumentation, 13, 595 - 606.

Steyn, D.W. (1977). Die konstruksie en evaluasie van 'n Suid-Afrikaanse Persoonlikheidsvraelys. Unpublished M.A. thesis, University of Pretoria, Pretoria.

Stout, R.L. (1981). New approaches to the design of computerized interviewing and testing systems. Behavior Research Methods & Instrumentation, 13, 436 - 442.

Sudman, S. & Bradburn, N.M. (1983). Asking Questions. A practical guide to questionnaire design. San Francisco: Jossey-Bass Publishers.

Sullivan, H.S. (1953). The Interpersonal Theory of Psychiatry. New York: Norton.

Sullivan, H.S. (1954). The Psychiatric Interview. London: Tavistock Publications.

Sundberg, N.D. (1977). Assessment of Persons. Englewood Cliffs: Prentice Hall.

Taylor, T.R. (1987). The future of cognitive assessment. HSRC special report PERS 420. Johannesburg: National Institute for Personnel Research.

Taylor, T.R., Gerber, N. & Rendall, M.E. (1981). The programming of psychological tests on the PLATO system. CSIR special report PERS 303. Johannesburg: National Institute for Personnel Research.

Tetlock, P.E. & Manstead, A.S.R. (1985). Impression management versus intrapsychic explanations in social psychology: a useful dichotomy? Psychological Review, 92(1), 59 - 77.

Thompson, K. (1984). Reflections on trusting trust. Communications of the A.C.M., 27, 761 - 763.

Tracey, T.J. (1985). Dominance and outcome: a sequential examination. Journal of Counseling Psychology, 32(1), 119 - 122.

Tracey, T.J. & Ray, P.B. (1984). Stages of successful time-limited counseling: an interactional examination. Journal of Counseling Psychology, 31(1), 13 - 27.

Tredoux, N. (1985). The development and programming of a computerized situation simulation for managers. HSRC special report PERS 345. Johannesburg: National Institute for Personnel Research.

Tuck, J.P. (1982). Will the real Achievement Anxiety test please stand up: effects of removing buffer items and altering item format of the Alpert-Haber achievement anxiety test. Psychological Reports, 51, 471 - 478.

Vale, C.D. (1981). Design and implementation of a microcomputer-based adaptive testing system. Behavior Research Methods & Instrumentation, 13, 399 - 406.

Wagman, M. (1982). Solving dilemmas by computer or counselor. Psychological Reports, 50, 127 - 135.

Warner, R.M., Kenny, D.A. & Stoto, M. (1979). A new round robin analysis of variance for social interaction data. Journal of Personality and Social Psychology, 37, 1742 - 1757.

Watts, K., Baddeley, A. & Williams, M. (1982). Automated tailored testing using Raven's Matrices and the Mill Hill vocabulary tests: a comparison with manual administration. International Journal of Man-Machine Studies, 17, 331 - 344.

Watzlawick, P., Weakland, J.H. & Fisch,R. (1974). Change. Principles of problem formation and problem resolution. New York: Norton.

Weiss, D.J. (1982). Improving measurement quality and efficiency with adaptive testing. Applied Psychological Measurement, 6, 473 - 492.

Weiss, R.L. & Moos, R.H. (1965). Response biases in the MMPI: a sequential analysis. Psychological Bulletin, 63, 403 - 409.

Wiggens, J.S.A. (1979). A psychological taxonomy of trait-descriptive terms: The interpersonal domain. Journal of Personality and Social Psychology, 37, 395 - 412.

Wright, T.L., Blackmer, D.R. & Ingraham, L.J. (1984). Simultaneous study of individual differences and relationship effects in attraction. Journal of Personality and Social Psychology, 47(5), 1059 - 1062.

Wright, T.L. & Ingraham, L.J. (1985). Simultaneous study of individual differences and relationship effects in social behaviour in groups. Journal of Personality and Social Psychology, 48(4), 1041 - 1047.

APPENDIX A

A run of items similar to those in the RAP. Options chosen by the test and the testee in a hypothetical testing session are marked with an asterisk.

ITEM 1

* Your next client is a Mr Dlamini. He says, loudly so that other clients can hear him: "I'm taking out all my money. Close my account, now!"

I -

1) say: "Yes, Sir. "

* 2) shrug my shoulders and say: "OK."

3) say: "Keep your voice down please."

4) smile and say: "So you've had a bad day, huh?"

ITEM 2

* Mr Dlamini shoves a transaction record book at you and points out where his salary cheque, which is always for the same amount, has been deposited every month for the past six months. It transpires that the most recent salary cheque was mistakenly put on hold by another teller the previous day. Mr Dlamini says "Its because you always think you know better."

I -

1) ignore him and take his record book to a superior for approval to have the cheque cleared.

2) first say: "If you can keep your mouth shut I'll have this fixed for you."

3) first say: "Relax - it's not so serious. We'll have this sorted out in no time."

* 4) first say: "I'm so sorry this happened. It's all our fault. I'll try and fix

ITEM 3

* When the cheque has been cleared and you ask Mr Dlamini how much he wants to withdraw he says: "I told you - close my account!"

When the cheque has been cleared and you ask Mr Dlamini how much he wants to withdraw he mutters: "Close my account, please."

When the cheque has been cleared and you ask Mr Dlamini how much he wants to withdraw he says, smiling weakly: "R100 please, although actually I want to close the account."

When the cheque has been cleared and you ask Mr Dlamini how much he wants to withdraw he says: "So you fixed it. That's good - I see you learn quickly. But I think I should close my account anyway."

I say -

1) "Whatever you want."

2) "Good riddance to bad rubbish."

* 3) "I'm sorry you're still upset. Can I do something to make you change your mind.?"

4) "Are you always like this?" and smile.

ITEM 4

At that point one of your superiors steps in and after some further discussion the account is finally closed on Mr Dlamini's insistence. You pay out the remaining balance in cash. After serving a number of other clients, you look up to find that Mr Dlamini is back. This time he holds up two stained and crumpled bank notes. He asks if you think this is money you paid him with.

At that point one of your superiors steps in and after some further discussion the account is finally closed. You pay out the remaining balance in cash. After serving a number of other clients, you look up to find that Mr Dlamini is back. He mutters something. When you ask him to repeat, he says you paid him dirty money and holds out two crumpled and stained bank notes.

* At that point one of your superiors steps in and after some further discussion Mr Dlamini agrees, as he puts it, to give you another chance. He withdraws R100 and you pay out the amount in cash. After serving a number of other clients, you look up to find that Mr Dlamini is back. He smiles broadly and says he thinks he's going to have to teach you how to do your job. He hands you two stained and crumpled bank notes and says you gave him those.

At that point one of your superiors steps in and after some further discussion Mr Dlamini agrees not to close the account. He withdraws R100 and you pay out the amount in cash. After serving a number of other clients, you look up to find that Mr Dlamini is back. He apologizes for bothering you again but says you gave him these, he holds out two stained and crumpled bank notes, and if you would mind exchanging them.

I exchange the notes -

1) but say: "We don't keep dirty notes. I didn't give you those."

2) and say: "Next please."

3) and say with a smile: "You again!"

* 4) and say: "Sorry about that."

ITEM 5

It is your lunch break shortly after this and as it happens you walk past Mr Dlamini in the supermarket. He looks at you in a pained manner.

It is your lunch break shortly after this and as it happens you walk past Mr Dlamini in the supermarket. He carefully avoids noticing you.

* It is your lunch break shortly after this and as it happens you walk past Mr Dlamini in the supermarket. He says "Hi!" and walks on.

It is your lunch break shortly after this and as it happens you walk past Mr Dlamini in the supermarket. He smiles.

In walking past I -

1) try to stare him into submission.

* 2) say: "Hi! How are you?"

3) say: "Hi! I hope you aren't angry anymore."

4) ignore him.

HSRC PUBLICATION LIST

A complete list of HSRC publications or a list of publications of a particular Institute of the HSRC is available from the Head: Technical Services

HUMAN SCIENCES RESEARCH COUNCIL
RAAD VIR GEESTESWETENSKAPLIKE NAVORSING

Functions of the HSRC

The HSRC undertakes, promotes, supports and co-ordinates research in the field of the human sciences. It also determines research priorities, disseminates the findings of human sciences research, facilitates and evaluates the implementation of research findings, stimulates the training of researchers, places the full spectrum of human sciences disciplines at the service of the inhabitants of the RSA and promotes science in general.

Institutes

Institute for Communication Research (ICOMM)

Institute for Educational Research (IER)

Institute for Historical Research (IHR)

Institute for Manpower Research (IMAN)

National Institute for Personnel Research (NIPR)

Institute for Psychological and Edumetric Research (IPER)

Institute for Research Development (IRD)

Institute for Research into Language and The Arts (IRLA)

Institute for Sociological and Demographic Research (ISODEM)

Institute for Statistical Research (ISR)

Bureau for Research Support Services (BRSS)

Administration

Head office

Private Bag X41, Pretoria 0001
Republic of South Africa
Telegrams RAGEN
Tel. (012) 202-9111
Telex 3-20893 SA

NIPR

P.O. Box 32410, Braamfontein 2017
Republic of South Africa
Telegrams NAVORSPERS
Tel. (011) 339-4451
Telex 4-25459 SA

Regional offices

Western Cape, Private Bag X5, Roggebaai 8012
Tel. (021) 419-2572/3/4/5 Telex 5-22260 SA

Natal, P.O. Box 17302, Congella 4013
Tel. (031) 815970 Telex 6-28567 SA

NIPR Eastern Cape, P.O. Box 1124, Port Elizabeth 6000
Tel. (041) 53-2131 Telex 2-43203 SA

Funksies van die RGN

Die RGN onderneem, bevorder, ondersteun en koördineer navorsing op die gebied van die geesteswetenskappe, bepaal navorsingsprioriteite, versprei die resultate van geesteswetenskaplike navorsing, vergemaklik en evalueer die implementering van die resultate van navorsing, stimuleer die opleiding van navorsers, stel die volle spektrum van dissiplines in die geesteswetenskappe ten diens van die inwoners van die RSA en bevorder die wetenskap in die breë.

Institute

Instituut vir Geskiedenisnavorsing (IGN)

Instituut vir Kommunikasienavorsing (IKOMM)

Instituut vir Mannekragnavorsing (IMAN)

Instituut vir Navorsingsontwikkeling (INO)

Instituut vir Opvoedkundige Navorsing (ION)

Nasionale Instituut vir Personeelnavorsing (NIPN)

Instituut vir Psigologiese en Edumetriese Navorsing (IPEN)

Instituut vir Sosiologiese en Demografiese Navorsing (ISODEM)

Instituut vir Statistiese Navorsing (ISN)

Instituut vir Taal- en Kunstenavorsing (INTAK)

Buro vir Ondersteunende Navorsingsdienste (BOND)

Administrasie

Hoofkantoor

Privaatsak X41, Pretoria 0001
Republiek van Suid-Afrika
Telegramme RAGEN
Tel. (012) 202-9111
Teleks 3-20893 SA

NIPN

Posbus 32410, Braamfontein 2017
Republiek van Suid-Afrika
Telegramme NAVORSPERS
Tel. (011) 339-4451
Teleks 4-25459 SA

Streekkantore

Wes-Kaap, Privaatsak X5, Roggebaai 8012
Tel. (021) 419-2572/3/4/5 Teleks 5-22260 SA

Natal, Posbus 17302, Congella 4013
Tel. (031) 815970 Teleks 6-28567 SA

NIPN Oos-Kaap, Posbus 1124, Port Elizabeth 6000
Tel. (041) 53-2131 Teleks 2-43203 SA

ISBN 0 7969 0540 1

1271/87*9