



1495(Dup)

# **Methodology issues in personality assessment research**

1495

paper presented by

**Cas H. Prinsloo**

on 9 June 2000

at the Industrial Psychology Conference  
(incorporating the Psychometrics Conference)  
Pretoria  
8&9 June 2000

## **Introductory remarks**

Two burning societal issues have often been raised recently in public debates and as part of policy development, not only in South Africa, but also abroad. Both of them have significance for applied personality assessment research. Both have as bedrock the phenomenon of escalating violence. The two examples both illustrate that personality assessment is considered for its potential utility, if not necessity, for screening people who are prospective carriers of firearms. In one case, the question has been raised in conjunction with the preparation of a new bill on the licensing of personal firearms. More specific, it has been suggested that people should be screened for suitability, personality wise, to possess a firearm. The other case is about whether or not personality testing would detect whether prospective carriers of weapons, as part of their task in security and other uniformed forces, would actually in future be likely to go rampant in the face of severe provocation or detrimental personal circumstances, for example.

These examples are raised, without going into them further in any depth, to illustrate the importance in daily life of understanding the profound theoretical and practical issues (of life and death, for that matter), which are at stake when considering the relevance and usefulness of personality assessment and personality assessment research.

More directly put, the present paper is a critical reflection on various issues of methodology when research is done to develop and improve personality assessment technology. As such, it also serves to provide a backdrop for the paper by Meiring (2000), in which recent personality assessment data are interpreted and used as illustration, so to speak, of many of the matters raised in this paper.

In order to enhance the authenticity of the duo of presentations, the layout or format of a typical empirical research article, as would be published in a scientific journal, is followed. It is hoped that this choice will enhance the perceived reality surrounding the issues raised.

A final observation made at the outset is that a deliberate choice was made to reflect only on very recent material. The rationale behind this choice is to acknowledge the "new debate" (which seems to have started) about personality tests, and advance it even more rapidly. Therefore, almost all sources referenced were published during the most recent five years. This choice was also exercised in full acknowledgement of the fact that many classic works on personality assessment research and methodology exist, dating as far back as 50 years, by authors as illustrious as Cattell, Eysenck and their peers. The publication by Mischel (1968) is a good case in point. Readers are reminded that a sufficient review of many of the fundamental issues raised today is impossible without consulting those authors and scholars.

## **On treating the background to the research**

In the "background" section to a research article or report, the researcher has to come to terms with certain customary items. These mainly are the existing body of accepted knowledge in the field, and previous findings reported in the literature. Thus, when doing personality assessment research, present-day scholars should demonstrate that they have imbedded their work in accepted research practice and a solid knowledge of the field.

It goes without saying that one such very important angle on the issue of personality assessment research is the one posed by the requirements of new legislation. Specific mention can be made of the *Employment Equity Act* (especially Section 8 of Act 55 of 1998), the *Labour Relations Act*, and the *Health Professions Act* (especially Section 37 (2) of Act 89 of 1997). Researchers in the field of test development and personality assessment should always demonstrate that they are *au fait* with such legal frameworks, and also elaborate on the implications of these recent acts and the ensuing debates. Mauer (2000) recently produced quite a helpful set of guidelines in this regard.

Another perspective that has to be dealt with is how present personality assessment research is linked with other related systems. As a result, networking through both informal contacts and forums, and formal scientific articles, is paramount. On the local scene, initiatives founded and statements released by trade unions, professional societies and bodies such as the Psychological Assessment Initiative, the National Productivity Institute and forums for personnel practitioners, are all very important. The thrust of all these activities is to account for relevant contributions by always being aware of gaps in our understanding (e.g., unsolved questions), and by trying to solve anomalies (e.g., contradictory research findings).

Two related matters also warrant emphasis. The first one is the principles and best practices at stake when importing Euro-American technologies (or any others, for that matter). Conscious, informed decisions have to be made on the appropriateness of doing so. However, there is no single and plain principle in this regard. The simultaneous justification of both the technology developer's intended usage, and the prospective user's application, is crucial. Sufficient overlap has to be found between the intended uses of developers and users with regard to the purpose and context of application and constructs. Test users and researchers should not deviate dramatically from the original scope and focus of the personality assessment technologies concerned, unless they thoroughly research and document their rationale for and success in doing so. By the way, this requirement of overlap, and many other stipulations, are "prescribed" amply at different places in the recent AERA/APA/NCME (1999) standards.

The second of the two matters raised in the previous paragraph, is how the relationship between theory and practice is perceived and treated. Does the researcher explore theory and try to explain new phenomena introduced by observations of reality, or does (s)he attempt to find confirmation in reality of theory-based predictions. Personality assessment research could clearly lean more towards either end of a continuum such as the one just mentioned, or indeed hover somewhere in between. However, researchers and test developers should know where they stand. They should also motivate and communicate any given stance clearly. Sometimes more theory-based points of departure, and the replication of existing positions through confirmatory studies are preferable. At other occasions pure empirical, *tabula rasa*, exploratory approaches are indicated, and new theory is actually being built.

As a result of the arguments raised so far, the methodology issues at stake in doing research on the development of assessment technologies for personality constructs are legion. One has to acknowledge and account for the foregoing aspects as well as those mentioned hereafter, amongst others.

Personality assessment is different from evaluating cognitive ability, aptitude, attitude, learning potential, interest, and other constructs. As a result, the former inevitably requires some unique approaches. To list a few instances, one can think of aspects such as item

formats, response sets, the notions of right / wrong and better / poorer in interpreting test "performance", test anxiety, test administration, time limits, and (de)briefing.

Within personality assessment research, various approaches exist. In fact, there are at least as many of them as there are personality theories. Each approach is valid and useful in its own right, provided that an appropriate fit is found between practical application requirements and the theoretical and conceptual scope of the relevant technique. Recent textbooks, such as the ones by Burger (1997), Gregory (1996), Groth-Marnat (1997), Handler and Hilsenroth (1998), Hogan, Johnson and Briggs (1997), Jeanneret and Silzer (1998), Mertens, (1998), Meyer, Moore and Viljoen (1989), Strickland (1998), Terre Blanche and Durrheim (1999), and Winter (1996) amply demonstrate the argument in point by illustrating in a variety of ways how personality theory and personality assessment are related. Personality is typically perceived as and measured from the viewpoint of being traits, developmentally determined, formed by social circumstances, biologically determined, and dynamic.

A facet often neglected, is that the level of assessment can legitimately range from very broad to very specific. One only has to look at the large volume of very recent publications, especially articles in scientific journals, on the assessment of the so-called Big Five factors. [No attempt is made in this paper to even start reflecting on them.] Work of this kind comprises what can be termed macro-level assessment, or the assessment of super-factors. Within the very lively debate, a great degree of consensus actually prevails on the identity and conceptualisation of the five super-factors. However, the most serious drawbacks of assessing personality at this level are practical. On the one hand, one does not end up knowing a lot about the test taker. On the other hand, the prediction value of correlates is not all that helpful, as the constructs are so broad.

At the meso-level, approaches exist where the so-called second-order factor structures of instruments are uncovered and applied. Such technologies can normally be validated quite well in terms of the predictive value of their constructs. Sometimes the constraint of being too general in nature, still applies.

At the micro-level, very useful and quite detailed assessments can be made, provided that care is taken not to over-extend the profiles or scores, especially when applied for high-stakes assessment. The latter refers to exclusive uses such as when screening some applicants out and selecting some in during job selection. Therefore, extra care has to be taken to crosscheck interpretation hypotheses with information from additional sources in a rigorous way.

The dust has not settled yet on the matter of the normal distribution of scores. Scholars differ vehemently on the kind of assumptions that could be considered correct with regard to it. Mabry (1999) recently challenged the transfer of the acceptance of normal distribution from certain mostly physical characteristics, such as height or weight, to non-physical entities. Normal distribution in the first case, according to Mabry, is insufficient ground for assuming that constructs such as intelligence or performance have to adhere to normalcy too. In the not too distant past, Herrnstein and Murray's (1994) release of their book on the (in)famous bell curve generated a fairly heated debate. Critics and supporters were many. The work by Fischer, Hout, Jankowski, Lucas, Swidler and Voss (1996) is a good case in point. The fact remains: - every personality assessment researcher today has an opinion on how the variables they work with, distribute, and should be able to justify her/his opinion.

Last, there is the matter of the universalism of traits *vis-à-vis* local variations. Williams, Satterwhite and Saiz (1998), for instance, researched the constructs of individualism and collectivism across cultures. They also connected the two constructs to the importance of psychological traits across cultures. This example is mentioned so that readers can take note of the debate about the (in)variance of "truth" or "knowledge", and explore this issue further in their own work. Does anxiety (and extraversion, and self-acceptance, and best child-rearing principles, etc.) "mean" the same in Europe and in the East? What about in Africa? Does a construct play the same role, have the same importance, and fluctuate equivalently relative to other core constructs everywhere?

Enough to argue that practitioners and researchers have to know where in history and present time they stand with regard to existing theoretical and empirical knowledge / findings. Let us further pursue the more detailed implications for method with regard to personality assessment research.

### **On making crucial decisions about research method**

Before inspecting specific matters related to test takers, instruments and procedures in more detail, a few general remarks are made.

It is prudent to remind ourselves again that the scope of instrument development research and related activities is determined by the purpose of the project. If the intention is to find confirmation for existing theoretical positions, the assessment instruments will be subjected to different procedures than when unknown territory has to be explored.

It is conceivable that research could be focused on the performance of whole populations. In such a case, representation and generalisation are very important. However, the research could also have an interest in the performance of specific target groups. In such an event, instruments will be customized in unique ways, along with purposive sampling procedures, to solicit to optimal degrees behaviour or performance that could be highly idiosyncratic. Any position on this continuum could be perfectly legitimate in as far as research objectives go.

Many textbooks and other relevant material, for example the guidelines for the validation of assessment instruments in the workplace produced by the Society for Industrial Psychology (SIP, 1998) highlight the various types of evidence that could be submitted in support of the validity of new and existing assessment technologies. It has to be understood that the various types of evidence have lesser or greater importance depending on the kind of application one has in mind. Be it as it may, face, construct, content, and criterion-related validity can be investigated and documented with the aid of numerous techniques.. Readers can consult many valuable articles by authors such as Clark and Watson (1995) and Schmidt and Hunter (1998) for further illumination. [By the way, the latter article testifies to the high validity of certain personality measures (often ranked second and third just after general cognitive ability assessments and job simulations, for instance)]. Discriminant-convergent analysis, factor analysis, and establishing expert opinion are but a few of the predominant routes to follow. Discussions indicating the need for each, and how it is best applied, especially with reference to the situation under which it should occur, can be found in many textbooks (Anastasi & Urbina, 1997; Huysamen, 1996).

Quite a substantive debate is the one on reliability *vis-à-vis* internal consistency *vis-à-vis* practical usefulness. Some experts would argue for the highest possible coefficients (in

excess of 0,7), as measured with the Kuder-Richardson 20 formula, for instance. Others, however, would rather accept lower internal consistency coefficients (ranging from 0,5 to 0,7) for gains at the end of practical coverage. Clark and Watson (1995) plead for a position closer to the latter. [Also note the pointed distinction made by these two scholars between unidimensionality and internal consistency.] It may not be that important at all to know with great confidence that one measures quite reliably or consistently a construct so limited that it becomes useless. In this regard, correcting for various response formats is also important. Note, for example, the Fergusson correction made when moving from dichotomous to polytomous response formats (Van den Berg, 1982). Related to the aspects mentioned immediately before, are the corrections that have to be made to coefficients when one wants to minimize the effect that an individual item has on its own scale total. More recent voices have gone so far as to question whether reliability actually is a necessary condition for valid measurement (Mabry, 1999). (Many other kinds of reliability and validity are proposed in the same breath, such as naturally occurring reliability, ecological validity, instructional validity and systemic validity.)

The way in which a researcher gathers her/his data and the nature of the data are also inextricably linked. The researcher has to be clear about why certain types of data from certain sources are preferred above others. This activity assumes that researchers make well-informed decisions about the applicability of methods, statistical techniques, and the nature of their data. It is no mere coincidence that questionnaires, self-report techniques, observation, simulation exercises, peer (or supervisor or subordinate) ratings, or whatever are all possible routes followed in personality assessment research. However, there has to be a well-established integrity and coherence between all the elements of one's research methodology. It thus becomes evident increasingly that each reason or ground for assessment necessitates the application of a different set of criteria or specifications as to the best route to follow in the actual achievement of the research purpose. (A few more notes are made further on in the "instruments" section with regard to this matter.)

Suffice it to say at this point that the crux of the matter is that conscious, founded decisions should be made with regard to research purpose and route when it comes to deciding on target populations and samples, research instruments and procedures.

- ***On determining who will comprise the sample or be the respondents***

The most important motivation that the personality assessment researcher has to give with regard to research participants, is why the actual test takers have been included in the materialised sample. Are they the best possible set of people to represent the intended target population? What is the nature of the compromises between scientific requirements and practical constraints (such as finances, deadlines and capacity available) that the researchers have had to make? Which factors have caused compromises, and how have they been dealt with? It has to be noted that in an ideal world, given unlimited budgets, time and people, no trade-offs would be necessary.

As can be derived from the previous paragraph, sampling is not an "either / or" case. To achieve a completely random sample is very seldom possible. However, the inability to achieve such ideal representation does not summarily imply the absence of any representation. In a way, there is a continuum of acceptable alternatives in between, such as clustering and stratification, and settling for various degrees of purposive or convenient sampling. At some point, the research team has to establish the best eventuality possible, and argue whether representation of the target population is sufficient or unacceptable. The purpose and type of research will also interact with the nature of the compromises

between practical, scientific and financial considerations to give the researcher grounds and substantiations for his/her final decision.

Research on or with assessment tools can have various aims or objectives. These range from new instrument development, to restandardising, norming and validating the tools, to establishing retest-reliability. Every aim has a unique influence on the anticipated nature and scope of the use of the assessment tools, as also alluded to by Prinsloo elsewhere (1999, 2000). Some of these matters also particularly relate to the strength and legitimacy of generalisations based on the research concerned.

Sampling should also be done with the specific minimum requirements which are demanded by the assessment technique, in mind. Many aspects concerning the characteristics of test takers may be crucial. Such aspects could include language proficiency, test anxiety, age, education level, response sets (Conn & Rieke, 1994), motivation, race, gender, how high or low the stakes are, and the inclusive as against exclusive nature of the intended assessment.

In summary, the most important point made is that sufficient overlap should exist between the circumstances and characteristics of test takers during the assessment research and the situation pertaining to the intended application later, in order to have any sense at all of the future quality, appropriateness and relevance of the particular assessment technologies under investigation or development.

- ***On selecting the instruments for the research***

A few remarks about the instrumentation are also appropriate.

The instrument format in itself is an important and very likely source of nuisance variance that has to be accounted for meticulously. Self-report, paper-and-pencil and computer-based versions each carries its own set of influences. These can manifest in different levels of anxiety, boredom, commitment and motivation among test takers. When the stakes of assessment with the tools being developed, are going to be high, additional conditions set in. These and other factors have to be carefully controlled, or monitored so as to be able to say something about the size of their effects.

The researchers have to pay attention thoroughly to two further important questions, which are posed next. How is the core instrument treated? The implication is that the personality instrument(s) under development can, amongst other things, be dealt with in a fairly isolated way (time- or history-wise, or in relation to parallel matters being measured), or be imbedded in a rich and wide context. How are the constructs operationalised? The implied matters concerned are aspects such as whether the existence of elements of personality are assumed from responses to preference statements, (dis)likes for samples of behaviour, value judgements, direct or indirect stimuli, etc.

Which supportive biographical and other empirical evidence is gathered to inform the development of the personality assessment instrument? Does these data allow meaningful investigations that will lead to conclusions about the conditions under which the test does or does not perform consistently, or as predicted, for instance?

The test data allow higher or lower degrees of sophistication in terms of analyses in relation to the kind of situations just noted, in relation to the investigation of discriminant and convergent validity, or multi-trait multi-method designs, for example. These would

allow for documenting whether or not the target / focus instrument performs as expected. Are differences and agreements between variables from different instruments theory consistent? Are there explanations for deviations from the normal distribution or other known and accepted distributions? Ellis and Mead (1998) provide excellent documentation about how the measurement equivalence of American and Spanish versions of the 16PF(5)® was investigated. Their findings and others of their kind also underpin present research done by the author to progress from the existing SA92 version of the 16PF in use in South Africa, to a South African 16PF(5)® equivalent. In that process, attention is given to many aspects such as construct refinement, language equivalence, readability (assessed by means of three different indices of proficiency), etc. Much more research is also required on matters such as the relative influence of home language, language of testing, and culture, not to forget the heritability or not of certain traits *vis-à-vis* others.

The whole issue of test imports, as was hinted at earlier, is a very involved one. Is transplanting Euro-American assessment technologies onto other continents per definition an unpardonable sin, or is there more to it? Making sense of one's answer to this question requires a thorough weighing of many arguments for and against. There are certain very obvious overlaps between the people living in various countries and cultures. These can be found especially in their shared experiences of and exposure to globalisation (if not certain activities *per se*, then their effects), electronic communication, urbanisation, market economies, science and technology, etc. However, differences can be equally salient in the predominance of rural lifestyles amongst some peoples, low literacy levels, and the general uniqueness of some populations.

- ***On selecting the best research procedures***

In addition to the many issues already listed at various places above, there are still a couple of matters worthy of comment. They pertain to the various procedures relating to test / questionnaire development as a process, including the use made of psychometrics. This section attempts to highlight that there is again a continuum of activities or approaches lying between two extreme ends.

At the one end, there exists the notion that personality test development is very standard (or standardised) in terms of dynamics and process. Here we can refer to the conventional, run-of-the-mill, general approach to the development and use of assessment tools with "mainstream" applications in mind. Not only could the whole approach be in view, but also selected elements. An example would be instruments for the screening of candidates in terms of general intellectual ability or even more specific aptitudes for the sake of deciding about their suitability for a career or job, for instance. Such instruments will have gone through broad standardization and norming for local conditions to allow comparisons between the individual's performance and that of a reference population. Validation and the establishment of reliability will have been core elements.

At the other end would be the relatively non-unitary and non-homogeneous dynamics of process where assessment tools have to be developed in ingenious ways to suit very unique purposes. Thorough exploration of both expected and unexpected elements is required. This kind of assessment research is typically custom-made by means of survey-type investigations and explorations (data gathering). A typical example would be the one-off development of an assessment model and subsequent instrument development to find and assess the ideal personality profile of a select few high-earning, technology specialists working under some extreme conditions without routine or structure on some secret job.



That there is (or should be) only one clear method and a one-and-only correct route towards the development of and research on every personality assessment tool, is a myth. That creativity and flexibility have to be balanced with sound procedures, in every unique case, is the imperative.

### **On reporting the research findings**

How findings are reported and communicated, on the one hand, includes the typical final scientific overview of research findings. These findings exist in the more conclusive and final versions of peer-reviewed research articles and the complete range of published test material in the form of user and technical manuals, questionnaires, answer sheets and scoring keys.

Work in progress, on the other hand, are more often published in a preliminary way to advisory committees, as internal or office reports, as provisional releases of assessment material, as conference papers, etc.

Clearly the focus and finality of these two types of products, and all the shades in between, comprise a variety of types of communication on research. As such, each unique report has to be interpreted for its own status and intention.

As a prominent special focus nowadays, the matter of cross-cultural equivalence has to be covered as well. It is again argued that the nature of our knowledge is incremental and imperfect. There are no absolute or perfect solutions. The continuum of the influence of culture → home language → test language on personality assessment has to be understood and accounted for. Recently, debate on this topic increased in special editions of journals (e.g., European Review of Applied Psychology, Volume 47(4) of 1997), new journals dedicated to the field, and textbooks. Many of the publications already cited earlier in the paper testify to this. Two additional examples to mention now would be the contributions by Triandis (1997), and Van de Vijver and Hambleton (1996).

### **On discussing the research findings**

Relating to the implications and applications side of the research findings, the interface between the test user and the test producer is briefly referred to again (Prinsloo, 1999).

Reaching conclusions on the basis of assessment research and assessment outcomes themselves is much like hitting straight balls with curved sticks. Another analogy would be an expert taking good photographs with a defective light meter, purely because he is so in tune with the shortcomings of his (for that matter - dated) equipment (personal communication, B. Steenberg, ca. 1990). As long as one understands and declares the areas of finiteness (limitations), and acknowledges the blemishes (implying more risky applications), it is often better having flawed assessment instruments than nothing at all. It is often better to be left to known degrees of unreliability or validity, rather than to high, total or unknown degrees of subjectivity (or unwarranted optimism). These conditions are influenced in major ways by rater, day of testing, mood, skill, intention, commitment, situation, inference in terms of construct / content, or assumption regarding applicability / prediction / other relations).

The nature and degree of overlap between the intentions, aims, interests, and other related issues as perceived respectively from the sides of the researcher and the

practitioner, have to be explored by the parties concerned. The responsibilities of each party, codes of ethic / conduct, and general "maturity" in the field have to be discussed more too.

### **On coming to conclusions about the outcomes of the research**

In conclusion, this paper brought arguments against handling research knowledge at any stage as absolute or final truth. It should rather be taken as an incremental process.

A plea is also made that we should use what we have and understand, with reasonable certainty, with modesty and the necessary caveats, never expecting perfection at any stage, - not even in the "final" product. (Another title for this paper could have been, "Required: The angel Gabriel to develop the perfect personality test".)

Various of the conceptual issues introduced in this paper are illustrated by Meiring (2000) in his paper, which provides a practical illustration case in point.

To end off, it only remains to appeal: Be aware and alert, navigators of this journey (practitioners, researchers and assessment technology developers and importers), that knowing the banks gives confidence to sail full ahead.

SIPPaper.doc

### **References**

- AERA/APA/NCME. (1999). Standards for educational and psychological testing. Washington: American Educational Research Association (AERA), American Psychological Association (APA), National Council on Measurement in Education (NCME).
- Anastasi, A. & Urbina, S. (1997). Psychological testing (7<sup>th</sup> ed.). Upper Saddle River, N.J.: Prentice-Hall.
- Burger, J.M. (1997). Personality (4<sup>th</sup> ed.). Pacific Grove, Calif.: Brooks/Cole.
- Clark, L.A. & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. Psychological Assessment, 7, 309-319.
- Conn, S.R. & Rieke, M.L. (1994) The 16PF® Fifth Edition Technical Manual. Champaign, Ill.: Institute for Personality and Ability Testing (IPAT).
- Ellis, B.B. & Mead, A. (1998). Measurement equivalence of a 16PF translation: An IRT differential item and test functioning analysis. In N.K. Tanzer (Chair). Cross-cultural assessment and test adaptations. Symposium presented at the 24<sup>th</sup> International Congress of Applied Psychology, San Francisco, Calif., 9-14 August 1998.
- Fischer, C.S., Hout, M., Jankowski, M.S., Lucas, S.R., Swidler, A. & Voss, K. (1996). Inequality by design: Cracking the bell curve myth. Princeton, N.J.: Princeton University Press.
- Gregory, R.J. (1996). Psychological testing: History, principles and applications (2<sup>nd</sup> ed.). Boston: Allyn & Bacon.

- Groth-Marnat, G. (1997). Handbook of psychological assessment (3<sup>rd</sup> ed.). New York: John Wiley & Sons.
- Handler, L. & Hilsenroth, M.J. (Eds). (1998). Teaching and learning personality assessment. Mahwah, N.J.: Lawrence Erlbaum Associates.
- Herrnstein, R.J. & Murray, C. (1994). The bell curve: Intelligence and class structure in American life. New York: Free Press.
- Hogan, R., Johnson, J. & Briggs, S. (Eds). (1997). Handbook of personality psychology. San Diego: Academic Press.
- Huysamen, G.K. (1996). Psychological measurement: An introduction with South African examples (3<sup>rd</sup> ed.). Pretoria: Van Schaik.
- Jeanneret, R. & Silzer, R. (Eds). (1998). Individual psychological assessment: Predicting behavior in organizational settings (1<sup>st</sup> ed.). San Francisco: Jossey-Bass.
- Mabry, L. (1999). Portfolio plus: A critical guide to alternative assessment. Thousand Oaks, Calif.: Corwin Press.
- Mauer, K.F. (2000). Psychological test and other similar assessment device classification and use in South Africa. <http://www.pai.org.za/board.htm> (People Assessment Initiative), 6 June 2000.
- Meiring, D. (2000). Revisiting the cross-cultural comparability of the 16 Personality Factor Inventory (16PF) in the South African context. Paper delivered on 9 June 2000 at the Industrial Psychology Conference (incorporating the Psychometrics Conference), Pretoria, 8&9 June 2000.
- Mertens, D.M. (1998). Research methods in education and psychology: Integrating diversity with qualitative and quantitative approaches. Thousand Oaks, Calif.: Sage.
- Meyer, W.F., Moore, C. & Viljoen, H.G. (1989). Personality theories: From Freud to Frankl. Johannesburg: Lexicon.
- Mischel, W. (1968). Personality and assessment. New York: John Wiley and Sons.
- Prinsloo, C.H. (1999). Fair personality assessment and the interface between test user and test producer. Paper presented on 1 March 1999 and 23 June 1999 at the Conference on Compliant Assessment Techniques arranged by the Institute for International Research, Midrand, 1-3 March 1999 and Sandton, 23-25 June 1999 respectively.
- Prinsloo, C.H. (2000, in preparation / ready for submission). Personality assessment research and instrument development in a changing policy environment: the case of the 16PF. Journal to be finalised, vol, pages.
- Schmidt, F.L. & Hunter, J.E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. Psychological Bulletin, **124**, 262-274.

- SIP. (1998). Guidelines for the validation and use of assessment procedures for the workplace. Johannesburg: Society for Industrial Psychology (SIP).
- Strickland, K. (1998). Reflections on assessment: Its purposes, methods, and effects on learning. Portsmouth, N.H.: Boynton/Cook. [Editor, Peter R. Stillman]
- Terre Blanche, M. & Durrheim, K. (Eds). (1999). Research in practice: Applied methods for the social sciences. Cape Town: University of Cape Town Press.
- Triandis, H.C. (1997). Cross-cultural perspectives on personality. In R. Hogan, J. Johnson & S. Briggs. (Eds). Handbook of personality psychology (pp. 439-464). San Diego: Academic Press.
- Van de Vijver, F. & Hambleton, R.K. (1996). Translating tests: Some practical guidelines. European Psychologist, 1, 89-99.
- Van den Berg, A.R. (1982). Betroubaarheid (Reliability). In A.R, van den Berg & J.F. Vorster, Basiese psigometrika (Basic psychometrics) (pp. 121-138). Pretoria: Human Sciences Research Council (IPER Training Series Nr 1).
- Williams, J.E., Satterwhite, R.C. & Saiz, J.L. (1998). The importance of psychological traits: A cross-cultural study. New York: Plenum. [The Plenum Series in Social/Clinical Psychology, C.R. Snyder (Ed.)]
- Winter, D. G. (1996). Personality: Analysis and interpretation of lives. New York: McGraw-Hill.