Use of the General Health Questionnaire (GHQ) in a Zulu-speaking setting: an assessment of translation, reliability and some validity issues.

Vaughn Mitchell John

Pietermaritzburg
January 1996

Declaration

This thesis was undertaken in the Department of Psychology, University of Natal, Pietermaritzburg and unless otherwise stated in the text, represents the author's own work. This thesis has not been submitted to any other university.

Vaughn Mitchell John
January 1996

ACKNOWLEDGEMENTS

Dedicated to the memory of a dear friend,
Sipho Moloko.
Hamba kahle Comrade.

ABSTRACT

Psychology in South Africa is facing a dire need for valid and reliable mental health instruments for all its citizenry. There presently exists a reliance on instruments of foreign origin. Very often such instruments are used without their psychometric properties having been tested in the local setting. The present study employed a multi-stage process for translating the General Health Questionnaire (GHQ) into Zulu. A simplified English version of the GHQ and the translated Zulu version were subsequently administered to a sample of two hundred and fifty seven (257) bilingual high school students. The data from this sample was used to assess the equivalency between the Zulu version and the English version. At the scale level, both versions of the GHQ showed adequate internal consistency and reliability. Item analysis revealed certain differences between the two versions. Possible explanations regarding semantic differences are discussed. Substantial overlap between the factor solutions of the two versions was found. These factor solutions were found to correspond well with those recorded in the literature. The present sample scored much higher on the GHQ than foreign samples do. Suggestions for raising the cutting-scores for South African samples are made. On the whole, the Zulu version displayed evidence of reasonable equivalence to the English version. A comprehensive research programme for the GHQ in South Africa is presented.

TABLE OF CONTENTS

SECTION A
Review of the Theory and Literature

SECTION B
Research and Discussion

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

"Health for all by the year 2000"

>   (World Health Organisation (WHO), 1978, cited in Seedat and
>   Nell, 1992, p185).

"Decades of apartheid policies have adversely affected the
provision of health and mental health care to many South
Africans"

>   (American Association for the Advancement of Science (AAAS),
>   1990, p5).

The first quotation describes the global programme of the WHO for
attaining a "state of complete physical, mental and social well-
being" (Seedat and Nell, 1992, p185). The second quotation is a
finding of the AAAS's Medical Mission of Inquiry to South Africa,
a foreign delegation that visited South Africa in 1989. The
juxtaposition of the above two quotes is significant, mainly
because the first outlines a global vision which South Africa
would need to subscribe to, and the second characterises the
challenge that this country would need to address in order to
turn the vision into reality.

South Africa is at an important time in its history. A period
that is characterised by transition, planning and development.
Most sectors and institutions of South African society are
experiencing some form of review and transformation. Not least of
these is the health sector. It is vital that psychology, in its
various forms of therapy, training and research, be included in
this transformation process. There is an urgent need to undo the
systemic and structural imbalances of the past so that psychology
can address the needs of this country and serve all its
citizenry.

Mental health services in the public sector are grossly inadequate and racially imbalanced. A very small proportion (10%) of registered clinical psychologists are employed in state posts. Thus, within the public sector, there is approximately one psychologist per 304 000 of the population. The same ratios for First World countries are in the region of one psychologist per 4000 (Kriegler, 1993). Interestingly, the apartheid health system has created a situation where the White population has a ratio that is even more favourable than First World countries. The ratio for White South Africans is approximately one psychologist per 3000 of the population. Given that 80% of the population is dependent on the public and welfare sector, there is a clear need for more practitioners in this sector to meet the country's needs (Kriegler, 1993).

Research in psychology is, arguably another of the key areas that is in need of change. On the whole, there is a paucity of baseline data about mental health (Miller and Swartz, 1992; Subedar, 1993). Miller and Swartz (1992) warn that the "planning of appropriate and accessible health services for all South Africans, within the economic constraints on such services, requires careful research into the distribution and frequency of ill health" (p52). Although reference is made to health in general, the advice of Miller and Swartz (1992) is also applicable to mental health and other areas where psychology can make a contribution.

Within research too, there exists racial inequities as well as an urban-rural imbalance. Researchers continually lament the absence of normative data for South Africans in general but for the African population in particular (Miller and Swartz, 1992; Viljoen, Levett, Tredoux and Anderson, 1994). Linked to this, are repeated complaints regarding the absence of appropriate psychological instruments for the African population (O'Neil, 1988; Subedar, 1993). O'Neil (1988) reports on the "difficulty of finding instruments translated into the vernacular and standardised on non-white South Africans" (p175).

Furthermore, it could be said that there is a general need for research and test development to be more sensitive to the multitude of cultural and linguistic diversities of the country.

The following pages contain a brief examination of the South African context and the state of mental health services.

Chapter 2

The South African context

In order to contextualise the present research, it is necessary to examine some of the conditions that currently prevail with regard to life in general in South Africa, but particularly with regard to mental health conditions. It is also necessary to examine aspects of the mental health system and to look at psychology's response to the prevailing conditions, especially in terms of the research conducted and the manner in which such research has been approached.

2.1 Health-related conditions in South Africa

Life in South Africa abounds with conditions that can be deemed to be stressful at the least and at many times, traumatic or debilitating. These are conditions that are all or mostly associated with a negative impact on mental health. In this context, mental health refers to the broader idea of mental well-being as advocated by the World Health Organisation (Seedat and Nell, 1992) and not to an absence of mental illness. With this in mind, a useful definition is, "Mental health refers to the optimal development of a person's potential and the effective coping with life-tasks and roles associated with a particular stage of life. Both personal and societal factors are involved in this process and interact with one another " (Gerdes, 1992, p41).

We share with other developing countries conditions such as high poverty levels, high unemployment, widespread illiteracy, a growing HIV-positive population, inadequate basic health care and generally poor socio-economic development. All of these conditions have a racial skew, affecting Africans the most.

The October Household Survey 1994 (Central Statistical Services, 1995), the first that was fully representative of the South African nation, reported on some of these conditions. Approximately 66% of the population earn less than R500 per month. The national average of unemployment is 33%. Within the 4.7 million unemployed, 4.1 million (88%) are African (Central Statistical Service, 1995). There are currently an estimated 1.5 million adults who have not received any formal schooling and a further 5 million people who could be considered to be functionally illiterate (Aitchison, 1995). Lindegger and Ward (1995) have stressed the considerable challenge that the HIV epidemic will place on psychology. The need for education, counselling and therapy for HIV and AIDS patients, and their families, will grow with the epidemic.

In addition to such typically Third World characteristics, there exists a plethora of conditions that are uniquely South African; a legacy of four decades of apartheid, approximately a decade of political violence, widespread and recurring industrial action, disproportionate access to resources and poor race relations. The birth of a democracy in the country has not erased the memory and effects of state oppression, hit squad killings and detention without trial. In this respect, the AAAS's Medical Mission of Inquiry to South Africa (1990), noted that "apartheid policies of the South African government have had a deleterious effect on the health of the majority of South Africans" (p19).

The diversities of its people is another distinguishing feature of this country. Of particular relevance to mental health services and this study, is the fact that South Africa is a multi-lingual society. There are eleven official languages and large parts of the society speak only a single mother-tongue language. In KwaZulu Natal the mother-tongue language spoken by the majority of its inhabitants is Zulu.

2.2 The response of psychology

2.2.1 Mental health services

The response to all the prevailing conditions that face the
country has been poor. Looking specifically at mental health
services, one can describe the provision as inaccessible,
inappropriate and inequitable (Pillay, 1987; Subedar, 1993;
American Association for the Advancement of Science, 1990).
Kriegler (1993) notes that "structural problems in mental health
services conjoined with attitudinal barriers are the cause of
inadequate mental health care in South Africa" (p64).

It is hoped that transforming the country's present system of
health care from a tertiary, curative model to a primary,
preventative one would create a more efficient and cost-effective
system. It is known that there are high rates of psychological
distress amongst working-class people who present at primary care
services (Miller and Swartz, 1992). It is therefore essential
that mental health professionals are part of primary health care
teams. It is also expected that future mental health care at
grass-roots level would be funded through a national health
insurance system (Kriegler, 1993). This should provide some
relief given that only 15% of Africans have access to medical aid
(Central Statistical Services, 1995). The present government has
allocated R500 million to the transition from tertiary to primary
health care in all of South Africa's nine provinces (Natal
Witness, 10 November 1995).

2.2.2 Mental health research

The response in terms of research has also generally been poor
and often inappropriate. More research, particularly
epidemiological studies, which gather baseline data about rates
and distributions of psychiatric disorders and psychological

problems will be required for an effective primary health care system.

However, attempts have been made to investigate some of the health-related conditions described earlier. There have been, albeit few, attempts to assess the impacts of such conditions on the mental health or the psychological functioning of citizens.

It must be reiterated that the inequalities within South African society has meant that many of the debilitating conditions have either been the sole experience of or have been concentrated in the lives of Africans; the group that constitutes the vast majority of South African citizens. According to the October Household Survey 1994, the total population for South Africa is 40.7 million. Of this, 30,9 million (76%) are Africans (Central Statistical Service, 1995). This majority sub-population has therefore become the group from which one would expect samples to be mostly drawn.

2.2.2.1. Research using foreign instruments

Studies conducted include some that have examined the effects of political violence (Carlyle, 1991; Turton, Straker & Moosa, 1991; Michelson, 1991, Liddell, Kvalsvig, Qotyana & Shabala, 1992; Pillay, Magwaza & Petersen, 1992; Dawes, 1994), the effects of stress in occupational settings (O'Neil, 1988; Govender, 1995), the effects of detention (Perkel, 1988) and the effects of unemployment (Leeb, 1986). A noteworthy characteristic of all these studies is the use of foreign instruments.

Most of the studies reported on thus far involve examinations of stressful or negative effects. The use of foreign instruments is not however limited to such studies but applies to South African studies in general, including studies of more 'normal' mental health issues and psychological functioning. An examination of articles appearing in the South African Journal of Psychology and

of masters and doctoral theses undertaken at South African
universities attests to the widespread use of foreign instruments
in South African research.

Some examples of such studies are reviewed here. In a cross
cultural study of attitudes to mental retardation, Rawlins (1983)
used the Disability Social Distance Scale. McGarr (1985) used the
Leader Behaviour Description Questionnaire (LBDQ-XII) for
assessing the influence of race on the perceived leadership role
of the first-line supervisor. The metamemory interview schedule
of Kreutzer et al. was utilized by Sharratt and van der Heuvel
(1995) to investigate metamemorial knowledge in a group of
African school children. van Ede (1995) examined the
applicability of the Metamemory in Adult (MIA) Questionnaire for
the South African population. Viljoen et al. (1994) used the
Bender Gestalt for establishing normative data for Zulu-speaking
children. Brand, Noordwyk and Hanekom (1994) examined the
effectiveness of using the Self-Directed Search (SDS)
questionnaire with respect to a group of African Std. 10 pupils.

It has already been noted that a common denominator in most of
the studies referred to earlier and of psychological
investigations in general, is the use of instruments of foreign
origin. The main focus of such studies however, is usually on
examining the relationships between situational variables and
conditions such as violence, unemployment and similar social
issues, and measures of mental health or functioning. The
instruments employed are thus a means to this end, that is,
providing indices of mental health or psychological functioning.

2.2.2.2. Translation of foreign instruments

The instruments employed have usually been developed in the
English language. The step of translation of instruments has
therefore become an almost standard part of the methodology
section of studies conducted with African research participants.

Unfortunately, because the objectives of the study are usually broader, the step of translating the instrument becomes a small part of the design and often receives inadequate attention.

This is clearly the case if one examines the methodology sections of most of the studies referred to earlier. As described in a later section, there now exists very detailed guidelines and effective methods for the translation of mental health status instruments, usually involving several steps and a variety of techniques (Brislin, 1970, 1980, 1986; Werner and Campbell, 1970; Sartorius and Kuyken, 1994; Sechrest, Fay and Zaidi, 1972). However, looking at the group of studies reported here, the attention given to translation and assessing psychometric equivalence of the instrument is scant and without much consideration of the available methods.

In the Rawlins (1983) study for example, the only details regarding translation given are that "the Disability Social Distance Scale was translated into Afrikaans by the faculty of Afrikaans and Nederlands at a White university" (p.113). With regard to the use of the scale with Zulu-speaking high school pupils, the author simply mentions the use of an interpreter who "clarified all the concepts for them in Zulu" (p.114). It is not surprising then, that one of the concluding statements of this study was, "Whether the instrument was suited to all cultures is a moot point" (p.159).

In the McGarr (1985) study the translation procedure employed with the Leader Behavior Description Questionnaire was a simple translation into Zulu by a person fluent in both English and Zulu and the translation was then checked by a university Zulu lecturer.

The translation procedures used in the Sharratt and van der Heuvel study (1995) and the van Ede (1995) study paid better attention to translation fidelity. Both studies employed the back-translation method for checking translation adequacy and

both involved a step of decentering or rewording of the original instrument.

While the latter two studies are improvements compared to the two former examples, they are certainly not exhaustive of the available methods for ensuring psychometric equivalence. Triandis, Vassiliou, Vassiliou, Tanaka and Shanmugam (1972) caution that "back-translation, which is so widely used in cross-cultural research, is not foolproof" (p.45). All four studies also display the common reliance on bilinguals in translation procedures. It is well-recorded that bilinguals use a language in a different manner to monolinguals using the same language. (Tyson, Doctor and Mentis, 1988; Church, Katigbak and Castaneda, 1988; Yang and Bond, 1980, Bond and Yang, 1982; Triandis et al., 1972). Such problems are also discussed in a later section that examines South African studies that have employed the General Health Questionnaire.

One can then conclude that a notable feature of South African psychological research, is the reliance on instruments of foreign origin that are simply translated into local languages. A further notable feature, but with more serious implications, is the use of such instruments where their psychometric equivalence with the original has not been established. The results of such studies become questionable and discerning real findings from artifacts introduced by an inadequately translated instrument becomes difficult. A further consequence is that comparisons made with other studies, usually foreign, employing the original instrument are shaky at best and spurious at worst. Viljoen et al. (1994) provide good support for this point on the basis of their results with a Zulu-speaking group which "clearly indicated the danger of using foreign norms in that the performance of the population group investigated proved to be significantly different to the performance of the foreign normative sample (i.e. American children)" (p.150). They concluded that using the foreign norms of the Bender Gestalt test for Zulu-speaking children would be inappropriate and would result in inaccurate assessment information.

2.2.2.3. Original development of instruments

The use of locally developed and validated instruments would be
an ideal solution to such problems facing research and practice
in psychology in this country.

However, there is a lack of sufficient original test development.
Compounding the problem is the fact that even less work has been
done towards developing instruments for all cultural groups.
Taylor and Boeyens (1991) note that there are no personality
instruments in South Africa that have been developed on all
cultural groups, creating a situation where personnel
practitioners have been forced to apply the 'White' test.

Whatever original test development there is, is limited largely
to the work conducted in large well-resourced para-statal
research bodies such as the Human Sciences Research Council
(HSRC) and the National Institute for Personnel Research (NIPR).
Huysamen (1983) reviews a number of psychological and educational
instruments developed by such institutions for South Africans.
Unfortunately, the research programme of institutions such as the
HSRC has tended to centre on the construction of separate
instruments or scales for South Africa's different race groups.

This practice has been criticised for amongst other things,
reifying the concept of culture and has been described as part of
the policy of separate development (Nell, 1990). Taylor and
Boeyens (1991) note that as "the apartheid system crumbles, test
populations become progressively more multi-cultural" (p1).
Further support can be found in van Eeden and Visser (1992), who
state that "compelling reasons presently exist for using common
psychometric tests, not the least of which is that in many
situations individuals compete for selection at tertiary level,
for jobs in industry, and the like. Separate tests may also lead
to discriminatory practices, with the result that common
psychometric tests that are unbiased, in other words, tests that
do not benefit certain population groups at the expense of

others, seem to be the wisest" (p163). It has been said that one of the major challenges facing psychology in this country, is the development of common psychological instruments for all the population groups (Owen, 1991).

Some examples of original test development that has occurred outside the larger para-statal institutions, include the work of Bluen and Odesnik (1988) in the development of a 42-item Township Life Events Scale, Perkel's (1988) study on detention, Duckitt's (1991; 1993) work on developing the Subtle Racism Scale and the design of an instrument to measure change in attitudes of white and black children, reported on by Taylor, Fourie and Koorts (1995). There is also work currently in progress at the University of Durban Westville on the development of the South African Mental Health Indicators (SAMHI) (Pillay, 1995, personal correspondence).

Some of the possible reasons for the dire shortage of original test development in South Africa include the following factors;

1. Test development work of high quality is expensive and time-consuming.
2. Lack of resources within most institutions.
3. Little interest in research of such a technical psychometric nature.
4. Little encouragement in training institutions for this type of research to be undertaken as masters and doctoral dissertations, possible aided by the perception that such research does not contribute to theory-building but merely provides a research tool.

Despite the disincentives, interests in theory and cross-cultural applications, actually demand that test-development research be conducted. In the absence of a variety of locally-developed tests and where one wants to make cross-cultural comparisons, the requirements of selecting instruments of proven worth and cross-cultural applicability, of rigorous multi-stage translation, of

establishing psychometric equivalence and of validation, becomes imperative. This study is seen as the start of this important process with regard to employing the General Health Questionnaire (GHQ) with the South African Zulu-speaking population.

2.2.3. To be able to use the words 'South Africa' as a prefix

As discussed earlier a consequence of the paucity of locally developed instruments, is the reliance on foreign instruments.

While the reliance on a foreign instrument is an understandable consequence, researchers have tended to employ commonly used instruments in an isolated and individualist fashion. One therefore finds several researchers using the same instrument, like the GHQ, but each conducting their own translations and adaptations of the instrument. It is argued that this practice hinders the creation of an identifiable South African version of the instrument. This situation is best presented if one makes comparisons with other non-western countries that also make heavy use of foreign instruments. In China for instance there are identifiable Chinese versions of a number of popular instruments; Chinese Minnesota Multiphasic Personality Inventory (MMPI), Chinese State-Trait Anxiety Inventory (A-State and A-Trait), Chinese Beck Depression Inventory (BDI) and the Chinese Purpose in Life Questionnaire (PIL) (Shek, 1989; Chan and Chan, 1983). It is noteworthy that the GHQ has also gained some official status in China arguably because of the level of research that has examined its psychometric properties in the Chinese setting (Chan and Chan, 1983; Chan, 1985).

It is rare to find such equivalents of instrument versions in this country, where the term 'South African' can be added as a prefix to the name of foreign instruments. While South Africa is admittedly more culturally diverse than China is, it is argued that the key difference between the South African situation of test usage and that of countries like China, is that in the

latter country, more research is invested in the instrument
itself and not just in applications of the instrument. The
instrument itself is subjected to more research that assesses its
psychometric properties and applicability in the local setting.
There are signs of a definite research programme for popular
instruments. It is the latter mentioned programme that allows for
the country's name to be used confidently as a prefix to the
instrument's name.

2.2.4. A proposed research programme for instruments of foreign origin

In order to make meaningful inferences from scores acquired via
any psychological instrument, the instrument must be shown to
possess a range of necessary qualities.

There must be some reasonable assessment that the construct, be
it mental health, locus of control or some other psychological
construct under measure, does indeed exist in the population of
study. Furthermore, the language and other cultural factors of
the population of study must allow for such a construct to be
operationalised and thereby render itself open to measurement.
The instrument must be shown to be a true measure of such a
construct. Such validity must be measured against other proven
measures or indices and done so with more than one sample.
The instrument must be able to produce the same score on a
repeated measures test. It must be a reliable measure of the
construct. Such an instrument acquires the merit of psychometric
soundness. In the section on 'Post-development studies' (p.29) the
GHQ is shown to be such an instrument.

However, when such an instrument is transported and intended for
use in a site that is different from its place of origin, these
same qualities must be re-established in the new setting. In
addition, a series of steps for ensuring equivalence needs to be
undertaken. This process may be a time-consuming and expensive

one. It is however necessary, and deserves the same amount of
care and precision that was involved in the development of the
original instrument. This process is proposed here as a 'research
programme' that needs to be undertaken for any foreign instrument
that is used locally. The proposed research programme for the GHQ
is presented below and the present study is envisaged as the
beginnings of this programme.

1. *Selection of instruments*

Nell (1994) notes that SA is not a beggar at the world's door.
Instruments must therefore be chosen with care. A number of
criteria must guide the selection of instruments;

a) the qualities of the instrument must have been adequately
   demonstrated in the country of development.
b) the instrument should be embedded in a sizable body of
   literature.
c) an added benefit exists if the literature points to previous
   successful transportation of the instrument.
d) the instrument must be cost effective. Viljoen et al. (1994)
   explain cost effectiveness to mean that an instrument must be
   easy to administer, must be able to be scored for quantitative
   analysis and must not require expensive test materials.

It is important to state at this point that the GHQ satisfies all
the above conditions (see Chapter 3).

2. *Rigorous translation of instruments*

Translations of high fidelity are imperative. The quality of such
translations needs to be tested using a variety of techniques and
methods available and should not be solely dependent on the use
of bilinguals. As outlined in the methodology section, the GHQ
will be put through an intensive multi-stage translation process.

3. *Reliability studies*

The instrument must be shown to provide reliable measures in a variety of ways. The present study provides a number of indices of the reliability of the GHQ.

4. *Validity studies*

The instrument must also be shown to be a valid measure of the construct it was designed to measure. Validity must also be established in a variety of ways. The present study provides data on the factorial validity of the GHQ. Further research would be required to establish other validity indices.

5. *Establishing norms*

In commenting on the lack of normative data for South African population groups, Viljoen et al. (1994) note that there "appears to be a strong need to establish appropriate norms for each test in the setting in which it is to be used" (p.145). Establishing local norms for the GHQ would require considerable further research. Such research would however be greatly enhanced by the establishment of one version of the GHQ that researchers can rely on and could be use repeatedly. It is the aim of this study to provide such a version of the GHQ in Zulu.

Subjecting the GHQ (and other foreign instruments) to such a programme would reap benefits for researchers, practitioners and theorists, and indeed, for psychology as a whole. If the GHQ emerges from the programme as an appropriate, reliable and useful tool in the South African context, the researchers employing the GHQ will have more confidence in their findings and the conclusions they come to. Practitioners will have a cost-effective first-stage screening device. Theorists may find that the proven quality of the GHQ may also have epistemological

gains. Meehl (1978) notes that one of the reasons why psychology has been so slow at theory building, is because researchers can always dismiss their failure at hypothesis support by claiming uncertainty regarding their instruments, and nonequivalence of instruments across related studies.

2.2.5. Rationale of the present study

In motivating the merits of the present study it may be useful to briefly list the salient points made thus far;

1. South Africa has a number of conditions that negatively affect mental health and warrant investigation.

2. The health care system in South Africa is being transformed from a tertiary care model to a primary care one.

3. Planning mental health services and responding to the needs of the country require baseline data and epidemiological studies.

4. There is currently a paucity of data available on mental health.

5. There is a lack of suitable instruments for gathering mental health data.

6. The shortage of instruments is most acute with regard to African language versions.

7. There exists a reliance on the use of foreign instruments for data gathering.

8. The appropriateness of foreign instruments in use is largely unchecked.

9. When translated, the psychometric equivalence of such new language versions is largely untested.

10. There is a dire need for instruments of proven worth that will allow for reliable and valid measures of mental health.

It is within this context that the proposed research is undertaken. The aim of this study is to provide a well-translated Zulu version of the General Health Questionnaire and to assess its psychometric properties within this specific South African setting.

Chapter 3

The General Health Questionnaire (GHQ)

3.1 Introduction

This section deals with the General Health Questionnaire, a mental health instrument developed by Goldberg and published in 1972. It examines the contextual background to Goldberg's project, the developmental studies conducted by Goldberg, subsequent use of the original English version in community and other psychological research, use of translated versions of the GHQ in various parts of the world and finally, use of the GHQ in South African studies.

3.2 Background to the General Health Questionnaire

The development of the GHQ is located within a period marked by a growing interest in 'community psychiatry', a move towards providing clinical care to individuals and population groups in community settings rather than in psychiatric institutions. The move is related to advances made in the development of psychotropic drugs in the 1950's (Subedar, 1993) and is also closely associated with the growing interest in epidemiology.

Goldberg (1972) describes the context in which the GHQ was developed as one of growing interest amongst psychiatric epidemiologists to measure the prevalence of non-psychotic illness in the community. At the time, the major problem that epidemiologists were faced with, was the lack of a reliable screening test of acceptable validity, for the identification of persons with non-psychotic psychiatric illnesses.

Epidemiology is defined as:

"... the study of the distribution and determinants of diseases and injuries in human populations . ... concerned with the frequencies and types of illnesses and injuries in groups of people and with factors that influence their distribution" (Mausner and Bahn, cited in Subedar, 1993, p7).

It is noteworthy, that more than two decades later, Subedar (1993) notes that "doing psychiatric epidemiological research is more difficult ... [because] few internationally acceptable instruments/tools are available that can be used for screening purposes in the field of mental disorders" (p2). She also makes the important point that some of the available tools are not practical for use in South Africa. A host of other difficulties associated with conducting epidemiological research in psychology are discussed by Miller and Swartz (1992).

The GHQ was developed in an attempt to redress this situation of paucity regarding instruments. Goldberg (1972) states that the aims of his project was to:

> devise a self-administered questionnaire that would
> identify respondents with non-psychotic psychiatric illness,
> by assessing the severity of their psychiatric disturbance.
> The questionnaire had to be easy to administer, acceptable
> to respondents, fairly short, and objective in the sense
> that it did not require the person distributing it to make
> subjective assessments about the respondent (p.1).

It was intended that the GHQ would provide an estimate of the degree of psychiatric disturbance of any individual without having to rely on the varying diagnostic standards of individual clinicians and irrespective of whether the respondent or anyone in his/her environment considers the respondent to be ill.

Goldberg and Hillier (1979) suggest that the GHQ should be thought of "as comprising a set of questions which form a 'lowest common multiple' of symptoms which will be encountered in the various differentiated syndromes of mental disorder, consisting as it does of symptoms which best differentiate psychiatric patients as a general class from those who consider themselves to be well" (p139).

To achieve its aim of detecting non-psychotic psychiatric illness at the time it is completed, the GHQ assesses the way the respondent has felt, thought and behaved in the time leading up to the occasion on which it is completed. Respondents are asked to compare the extent of their current experience of each item with the extent to which it is usually experienced. The item is scored as being present only if it is being experienced 'more than usual'. This focuses the measurement away from long-term possession of neurotic traits. In this way, the GHQ focuses on two major classes of phenomena, namely, the inability to continue one's normal healthy functions and the emergence of new distressing phenomena (Layton and Rust, 1986).

For each item, a four-point response scale ranging from 'Not at all' to 'Much more than usual', is offered. As described by Goldberg (1972) and illustrated below, the scale may be scored in two possible ways, a likert score ranging from 0 to 3 or the binary method of 0 or 1, which is referred to as the GHQ score.

	Likert score	GHQ Score
Not at all	0	0
No more than usual	1	0
Rather more than usual	2	1
Much more than usual	3	1

Goldberg (1972) offered three potential uses for an instrument
such as the GHQ, which would assign an individual to a position
on an axis of mental health, irrespective of whether or not the
individual is thought to be a 'case' in the psychiatric sense.

1. Such an instrument would make it possible to compare the
amount of psychiatric disturbance in two populations by a
comparison of the means and standard deviations of scores in each
population.

2. With a longitudinal design, a given population could be tested
on different occasions in order to follow the changes in
psychiatric disturbance that occur with time.

3. Psychiatric disturbance as assessed by scores on the
questionnaire could be correlated with other clinical and social
variables in a given population (Goldberg, 1972, p3).

All of these are in essence fundamental goals within
epidemiology. As reported in subsequent sections of this chapter
it will be shown that the GHQ has been put to much broader use in
a wide variety of settings. Of particular relevance to the
present study, is the many cross-cultural applications of the
GHQ.

3.3 The model of illness and illness-detection implied in the GHQ

The GHQ is based on the conception of illness which presupposes
an axis of psychiatric disturbance on which any individual can be
placed. The axis ranges from severe disorder to a state of
hypothetical normality. A number of steps between these end-
points are envisaged to represent the various degrees of
disturbance.

Furthermore, psychiatric disturbance is thought of as being evenly distributed throughout the population in varying degrees of severity. The GHQ was therefore conceived of as an instrument that would detect persons who are ill by assessing the severity of their disturbance. When a given respondent is assigned a position on the axis according to known limits of error, the individual's score on the GHQ is thought of as a quantitative estimate of that individual's degree of non-psychotic disturbance (Goldberg, 1972).

The term 'neurotic' was deliberately avoided in Goldberg's work and replaced with 'non-psychotic', because of the contrasting ways in which psychiatrists and psychologists use the term 'neurotic' and because of the ill-defined manner in which medical doctors have applied the term.

It is important to emphasise that the GHQ was not developed as an alternative to existing procedures of illness detection via clinical interview. Goldberg (1972) notes that the "main value of a screening test in epidemiological research will be the first stage of what is essentially a two-stage process of case identification, the second stage being a full clinical interview by an experienced psychiatrist" (p3).

The distinction that is being drawn between screening and case-finding is an important one in mental health epidemiology. Case-finding, the second stage referred to by Goldberg, is a relatively well-established practice conducted regularly by trained clinical professionals and is commonly referred to as the 'clinical interview'.

Screening by contrast, as implied by Goldberg, is described as, "the presumptive identification of unrecognised disease or defect by the application of tests, examinations, or other procedures which can be applied rapidly to sort out apparently well persons who probably have a disease from those who probably do not. A screening test is not intended to be diagnostic." (Commission on

Chronic Illness, 1951, cited in Subedar, 1993, p.14.)

The relationship between screening and case-finding in a two-stage model of detecting mental disorder, is clarified by Blum (1962, cited in Goldberg, 1972), who spoke of 'potential cases', that is, those picked out by questionnaires, key informants, et cetera (screening) and 'actual cases', which are those confirmed by via psychiatric interview (case-finding).

3.3.1 The Goldberg and Huxley model

Goldberg and Huxley (1980) propose a hierarchical model of five levels (see Figure 3.1) to describe the nature and detection of mental illness in the community. An examination of the model here helps to establish an understanding of the role and value of an instrument such as the GHQ within a primary health care system.

Each of the levels in the model describes different populations of persons with mental illness and in order to pass from one level to the next, a person would have to pass through a filter.

Level 1 comprises the community with all psychiatric disorders present. Knowledge at this level is derived from surveys that have screened for psychiatric morbidity in the entire population or a random sample of it.

Level 2 comprises the patients that make up psychiatric morbidity within the primary care setting. The size of this group is not dependent on the physician actually detecting the illness. The first filter is between level 1 and level 2 and is referred to as 'illness behaviors'.

Level 3 comprises the attending patients who are identified as mentally ill by their doctor. This group collectively represent the psychiatric morbidity of the primary care setting. The filter between level 2 and level 3 is represented by the doctor's

ability to identify mental illness in his/her patients. The characteristics of both patient and doctor influence the passage through this filter.

Level 4 comprises those patients who present at psychiatric out-patient clinics and private practices. The filter between level 3 and level 4 is the doctor's referral of patients to psychiatric out-patients.

Level 5 comprises those patients that are admitted to mental hospitals. The filter between level 4 and level 5 is the decision made by a psychiatrist to admit.

Figure 3.1: The Goldberg and Huxley model.

```
┌─────────────────────────────────────────────┐
│   ┌─────────────────────────────────────┐    │
│   │            Level 1                  │    │
│   │  Psychiatric morbidity in the       │    │
│   │          community                  │    │
│   └─────────────────────────────────────┘    │
│                                               │
│        1st Filter: Illness behaviours         │
│                                               │
│   ┌─────────────────────────────────────┐    │
│   │            Level 2                  │    │
│   │ Total psychiatric morbidity within  │    │
│   │          primary care               │    │
│   └─────────────────────────────────────┘    │
│                                               │
│      2nd Filter: Identification by general    │
│                 practitioner                  │
│                                               │
│   ┌─────────────────────────────────────┐    │
│   │            Level 3                  │    │
│   │ Identified psychiatric morbidity    │    │
│   │       within primary care           │    │
│   └─────────────────────────────────────┘    │
│                                               │
│        3rd Filter: Referral by general        │
│                 practitioner                  │
│                                               │
│   ┌─────────────────────────────────────┐    │
│   │            Level 4                  │    │
│   │   Psychiatric out-patient attendees │    │
│   └─────────────────────────────────────┘    │
│                                               │
│      4th Filter: Decision to admit by         │
│                 psychiatrist                  │
│                                               │
│   ┌─────────────────────────────────────┐    │
│   │            Level 5                  │    │
│   │      Psychiatric in-patients        │    │
│   └─────────────────────────────────────┘    │
│                                               │
│  Source: Goldberg and Huxley (1980)           │
└─────────────────────────────────────────────┘
```

The Goldberg and Huxley (1980) model is useful for examining the pathways of detection and service delivery within a mental health system. The model will perhaps gain greater relevance as South Africa moves further along to the proposed primary health care system. At that point, an application of the model will highlight the role of an instrument such as the GHQ within level 1 and level 3. The GHQ could be used for screening for psychiatric morbidity within the community at level 1 and for assisting primary care personnel in detecting mental illness at level 3.

With reference to the value of instruments such as the GHQ at level 1, Goldberg and Huxley (1980) state that:

> these research instruments have been used to measure rates for psychiatric illness in the general population in order to arrive at estimates of prevalence independent of the illness behaviour of the patient or the ability of his medical attendants to detect and treat any disorder that may present. When this is done the concepts of psychiatric illness which have been derived from those patients seen by psychiatrists are being back-projected onto the general population in order to assess the numbers of those with similar patterns of symptoms who have not sought psychiatric care (p3).

The value of the instrument at level 3 is further explored in a later section (see p29) that deals with its use within the general practice setting. In this regard, it is known that large parts of the population present at primary care facilities with mental health problems. Some studies suggest that the number may be as high as between 50% and 70% of a primary physician's case-load (Kriegler, 1993). It is also known that the presence of psychological distress increases the consulting rates of people in the community to their general practitioner (Burvill and Knuiman, 1983). In many instances such patients are unaware of the psychological nature of their problems.

As utilised in other parts of the world (Skuse and Williams, 1984; Finlay-Jones and Murphy, 1979; Fontanesi, Gobetti, Zimmerman-Tansella and Tansella, 1985; Medina-Mora, Padilla, Campillo-Serrano, Mas, Ezban, Caraveo and Corona, 1983; Mari and Williams, 1984), the GHQ in the hands of primary care personnel could serve as a useful screening tool for assessing such patients. Skuse and Williams (1984) outline four practical strategies for how the assessment of the general practitioner, or other health care personnel for that matter, can be combined with the GHQ for screening purposes.

3.4 Goldberg's development and assessment studies

An assessment of the efforts put into the development of an instrument is crucial to any evaluation of that instrument. Such an assessment helps in determining the potential wider application and suitability of the instrument. The GHQ in its present form, emerged from an extensive research programme, involving rigorous design and large and varied samples of respondents.

According to Goldberg (1972), generating items for the GHQ was based on a review of research on symptomatology, examinations of existing instruments such as the Taylor's Manifest Anxiety Scale, Eysenck's Maudsley Personality Inventory and the Minnesota Multiphasic Personality Inventory (MMPI), and on the clinical experience of several psychiatrists. The search for items was guided by the decision to cover four main areas, namely, depression, 'felt psychological disturbance', objectively observable behaviour, and 'hypochondriasis'. The process resulted in the accumulation of 140 items, about equally divided between the four areas.

The 140 items constitute what is referred to as the "Long Form of the Questionnaire". The next step involved a calibration of the long form, involving three calibration groups, namely, 'normals',

'mildly ill' and 'severely ill'. Part of this stage involved the elimination of items which did not discriminate adequately between the three groups, allowing the establishment of the 60-item version. This item-reduction process led to the emergence of a single unidimensional scale, the 60-item version.

Once calibration was completed, the 60-item version was subjected to a range of reliability and validity studies, including

1) a test-retest reliability study involving 120 patients
2) a split-half reliability study based on 853 completed questionnaires
3) a validity study based on 200 general practice patients who completed the GHQ and a psychiatric interview
4) a validity study based on 91 patients in a medical out-patients department

The reliability and validity studies were also used to order the individual items in terms of how they performed in the various studies, thus allowing for the establishment of even shorter forms of the GHQ, namely, a 36-item, 30-item, 20-item and 12-item version. Full details and results of these studies are provided in Goldberg (1972). The latter work also contains reliability and validity data for all of the shorter versions and the cut-off scores for determining 'caseness'. All of the shorter forms of the GHQ correlate well with the 60-item version (Goldberg, 1972). On the basis of the GHQ literature reviewed for this study, the shorter 30-item version appears to be the one most extensively employed in research. The present study also employed the 30-item version.

A further shorter version was subsequently established by Goldberg and Hillier (1979). Referred to as 'a scaled version' this 28-item version of the GHQ consists of four subscales, namely, somatic symptoms, anxiety and insomnia, social dysfunction and severe depression. Only half of the items in the 28-item version appear in the 30-item version.

3.5 Post-development studies

3.5.1 Applications of the GHQ in research

As noted earlier in the section entitled, 'A proposed research
programme for instruments of foreign origin' (p.14), the choice of
a test should, amongst other things, be dependent on it having a
a sizable body of literature. The review of literature on the GHQ
for the present study is based on more than 40 studies that have
employed the GHQ. The literature on the GHQ is certainly
considerable given that the present literature search did not
make use of available computerised databases. Of noteworthy
importance, is that many of these studies have examined the
validity of the GHQ.

Studies have reported the successful employment of the GHQ in a
variety of settings as listed below;

As originally intended, the GHQ has been used as a screening
device for psychological distress in community settings
(Goldberg, Rickels, Downing and Hesbacher, 1976; Tarnopolsky,
Hand, Mclean, Roberts and Wiggins, 1979; Burvill and Knuiman,
1983; Stanley and Gibson, 1985; Goodchild and Duncan-Jones, 1985;
Cairns, Wilson, McClelland and Gillespie, 1989),

For similar screening purposes the GHQ has been used in general
practice settings (Skuse and Williams, 1984; Finlay-Jones and
Murphy, 1979; Vachon, Sheldon, Lancee, Lyall, Rogers and Freeman,
1982; Fontanesi et al., 1985; Medina-Mora et al., 1983; Mari and
Williams, 1984).

A further screening site has been hospital out-patient and clinic
settings (Mari and Williams, 1985; Vazquez-Barquero, Padierna
Acero, Marton, and Ochoteco, 1985; Lobo, Perez-Echeverria and
Artal, 1986; Ormel, Koeter and van der Brink, 1989; Gureje and
Obikoya, 1990; Aderibigbe and Gureje, 1992; Dhadphale, Ellison
and Griffin, 1983).

The GHQ has also been used for assessing psychological symptoms within certain student populations including, secondary school students (Shek, 1989), college students (Sriram, Chandrashekar, Isaac and Shanmugham, 1989), first-year university students (Chan and Chan, 1983; Surtees and Miller, 1990; Miller and Surtees, 1991) and with recent school-leavers (Banks, 1983).

Considerable success has been reported with the GHQ within occupational settings. Banks, Clegg, Jackson, Kemp, Stafford and Wall (1980) found the GHQ to be psychometrically sound in three studies employing large samples of employees in an engineering firm (n=659), recent school-leavers (n=647), and unemployed men (n=92).

Several studies have used the GHQ in investigating the mental health effects of unemployment. Within this field of interest, GHQ scores have been compared with variables such as sex (Winefield and Tiggemann, 1985; Banks and Jackson, 1982), age cohorts (Broomhall and Winefield, 1990) length of unemployment (Hepworth, 1980), lifestyles of the unemployed (Kilpatrick and Trew, 1985; Brenner and Bartell, 1983) , the moderating roles of employment commitment (Jackson, Stafford, Banks and Warr, 1983) and work involvement (Stafford, Jackson and Banks, 1980), and the effects of rehabilitation centres (Kemp and Mercer, 1983). Jenkins, MacDonald, Murray and Strathdee (1982) used the GHQ in examining the mental health effects of the threat of redundancy in a professional group.

Most of the studies on unemployment have used the shortest form of the GHQ, namely, the 12-item version. This was most likely because the GHQ formed part of a larger battery of questionnaires and/or scales. Given the high levels of unemployment in South Africa reported earlier (p5) such widespread and effective use of the GHQ with this population augurs well for similar use of the instrument in this country. Leeb (1986) has already put the GHQ to such use. A further demographic feature of this country also has bearing on this particular discussion; more than half the

South African population is made up of youth, younger than 16 years, many of whom are unemployed (Kriegler, 1993). In this regard it is significant to note that the GHQ has been successfully employed with young unemployed samples (Banks et al., 1980; Banks and Jackson, 1982; Broomhall and Winefield, 1990; Stafford et al., 1980, Jackson et al., 1983). The present study also employs a young sample.

3.5.2 Validity studies of the original English version

Many of the studies referred to above in terms of the different settings in which the GHQ has been applied have commented on the validity of the instrument.

Goldberg et al., (1976) compared the GHQ with the Symptom Checklist (SCL) and found both to correlate equally well against independent clinical assessment, particularly with symptoms of anxiety and depression. Banks (1983) validated the 30-item, 28-item and 12-item versions of the GHQ against the Present State Examination (PSE) and found all three versions of the GHQ to correlate highly with the PSE. Skuse and Williams (1984) compared the GHQ with the case-detecting behaviour of the general practitioner (GP) and an independent psychiatric assessment. In this sample where the estimated true prevalence of psychiatric 'cases' was 34%, the GP's classified only 24% as 'cases' while the GHQ classified 39% as cases. Mari and Williams (1985) validated the 12-item version of GHQ and the Self Reporting Questionnaire (SRQ-20) against the Clinical Interview schedule (CIS). The correlation between both instruments was +0.72.

3.5.3 Sensitivity and specificity and misclassification rates

Sensitivity is the proportion of cases that are correctly identified. Specificity is the proportion of non-cases that are correctly identified. The overall misclassification rate is the proportion of false negatives and false positives identified. All three are important validity indices. In the Mari and Williams (1985) study the GHQ showed higher sensitivity (85%) and lower specificity (79%) than the SRQ-20 (83% and 80%, respectively). The GHQ also had a better overall misclassification rate (18%) than the SRQ-20 (19%). Other workers have also reported acceptable levels of sensitivity and specificity for the GHQ (Banks, 1983; Goldberg et al., 1976). The highest validity coefficients of the GHQ were reported by Goldberg and Blackwell (1970) were a sensitivity of 96%, specificity of 88% and misclassification rate of 9% was found. The median values for sensitivity and specificity obtained from twelve published studies that employed the 28-item version is 86% and 82%, respectively (Aderibigbe and Gureje, 1992).

Studies that have reported somewhat lower agreement rates between case-detection by the GHQ and case-detection by a psychiatrist, are important for identifying reasons for misclassifications by the GHQ. False negatives on the GHQ are more likely to be people with chronic disorders, particularly anxiety states and particularly with women. Defensive individuals are also more likely to emerge as false negatives. False positives on the GHQ are likely to be people distressed by severe physical illness, a recent adverse life event, have major family burdens or loneliness (Finlay-Jones and Murphy, 1979; Stanley and Gibson (1985). False negatives pose a more serious problem in mental health screening than false positives.

In general the GHQ in its original English version and in its various translated version (discussed later, p36) has been shown to possess good validity as a first stage screening device.

3.5.4 The question of chronicity

The influence of chronic disorders on false negatives on the GHQ has generated some interest and attempts at finding solutions to this problem.

Goodchild and Duncan-Jones (1985) propose a revised scoring method for the GHQ, which they argue is a better prediction of caseness than the usual method of scoring. Their argument is based on the view that a response of 'no more than usual' to an item describing pathology, should be treated as an indicator of chronic illness rather than good health. They support their argument with evidence in which the revised scoring is associated with other measures of neurotic illness or trait neuroticism. The revised scoring is also shown to be more stable in repeated measurement.

In a study designed to check whether the validity of the GHQ is indeed improved by the rescoring procedure proposed by Goodchild and Duncan-Jones, Cairns et al., (1989) report no such improvements. Failing to replicate Goodchild and Duncan-Jone's (1985) findings, they note that there was "no evidence that the new scoring scheme improved correlations between the GHQ and other measures of morbidity nor did this (new scoring method) improve sensitivity or specificity when validated against the PSE (Present State Examination) " (p793).

In an attempt to decrease the number of false negatives in community studies, Stanley and Gibson (1985) recommend the use of a short questionnaire which detects the commonest false negatives, together with one of the shorter forms of the GHQ. In their study they designed a ten-item Mental Health Scale (MHS), which assesses chronic neurotic disorder. The MHS was combined with the 30-item GHQ.

3.5.5 Factor analytic studies

The development work on the GHQ favoured the emergence of a single unidimensional scale. Not surprisingly, subsequent attempts by Goldberg himself and other workers (Medina-Mora, 1983; Layton and Rust, 1986) were not greatly successful in constructing subscales from the component parts of the GHQ.

Layton and Rust (1986) conducted a factor analysis with the 60-item version of the GHQ in a group of male school children and in a group of men facing redundancy, using oblique rotations. The five factors that emerged for the school group were labelled, 'Depression and anxiety', 'Insomnia and anergia', 'Somatic symptoms', Difficulty coping' and 'Loss of confidence'. In total these five factors only accounted for 41% of the variance. The five factors for the group facing redundancy were labelled, 'Social dysfunction and loss of confidence', 'Severe depression', 'Insomnia and personal neglect', 'Somatic symptoms' and 'Anhedonia'. Together these five factors only accounted 33.2% of the variance. By contrast, Sang (1992) found three major factors in a study with Vietnamese refugees, which accounted for 75.1% of the total variance.

The Layton and Rust (1986) study clearly supported the idea of a single GHQ factor of mental well-being. Factors subsequent to the first factor were small, most of them with less than 6% variance for both samples. Factor analysis of the GHQ in a Mexican hospital general practice setting (Medina-Mora et al., 1983) also found factors subsequent to the first to be small; of 5.4% and lower.

3.5.6 The problem of retest effects

Much of the research using the GHQ has involved cross-sectional designs (Broomhall and Winefield, 1990; Mari and Williams, 1984; Gureje and Obikoya, 1990). A few studies have attempted longitudinal research (Ormel et al., 1989).

Ormel et al. (1989) cite a study by Henderson et al. which demonstrated a substantial retest effect for the GHQ. Ormel et al. (1989) themselves found substantial retest effects in a three-wave longitudinal study among new psychiatric out-patients. They propose that the retest effects in their study are best explained by social desirability and legitimation hypotheses. The 'social desirability hypothesis' attributes the retest effects to the respondents' desire to present themselves more favourably at subsequent measurements, thereby demonstrating an improvement which they see as expected of them. The 'legitimation hypothesis' suggests that the retest effects are due to deliberate exaggerations of symptoms at the time of first measurement. It proposes that respondents inflate their GHQ scores at initial measurement in order to engage the interest and sympathy of the health professional.

The existence of retest effects would affect the usefulness of the GHQ in longitudinal population and outcome studies. Given the repeated measurement employed in the design of the present study, retest effects will need to be considered in interpreting the data.

3.5.7 The Interval-General Health Questionnaire (I-GHQ)

An adaptation of the GHQ for studies involving repeated measures, has led to the establishment of the Interval General Health Questionnaire (I-GHQ) by Surtees and Miller (1990). The I-GHQ which is intended to assess more minor psychological conditions over a period of time, is a technique based on 12 selected items

from the 30-item GHQ combined with an adaptation of the Longitudinal Interval Follow-up Evaluation (LIFE) procedure (Miller and Surtees, 1991; Surtees and Tansella, 1990). The 12 items selected are not the same as those making up the 12-item version GHQ.

Miller and Surtees (1991) used the techniques to assess psychological symptoms and their course in first-year medical students. They report evidence which shows that the I-GHQ allied to the principle of LIFE, allows for a reliable and valid assessment of the course of minor spells of psychological symptoms.

With the I-GHQ each of the 12 items is assessed and scored in the usual way at intervals which may be several months. At the subsequent times of administration, changes in the item scores during the preceding period are probed, allowing respondents to use their initial and final scores as anchor points.

3.6 Translated versions of the GHQ

According to Goldberg and Williams (1988, cited in Sriram et al., 1989), the GHQ has been translated into not less than 36 languages. In India the GHQ has been translated into Hindi and Bengali. The 12-item version was successfully employed as a first stage screening tool in an investigation of psychiatric morbidity in Bangalore City (Sriram et al., 1989).

A Spanish version of the 60-item GHQ has been reported to work well (Mari and Williams, 1984). The 30-item version has been used in epidemiological studies in Spain (Lobo et al., 1986). The 60-item version was used by Vazquez-Barquero et al. (1985) in an investigation of the psychiatric correlates of coronary pathology.

In China, Shek (1989) investigated the validity of the Chinese version of the 30-item GHQ with secondary school students. Chan (1985) assessed the equivalence of the Chinese version with bilingual medical students, and Chan and Chan (1983) investigated the reliability, validity and structure of the GHQ with undergraduate students. Sang (1992) translated and validated a Vietnamese version of the GHQ with a refugee population.

Mari and Williams (1984), report that good results were obtained in a study that employed a Yugoslav version of the GHQ. In Mexico, the factor structure of the GHQ has been investigated with general hospital out-patients (Medina-Mora et al., 1983). In Brazil, a Portuguese version of the GHQ was used in a study of psychiatric morbidity in a primary care setting (Mari and Williams, 1984). Fontanesi et al. (1985) validated the Italian version of the 30-item version in a general practice setting.

As with the original English version, validity studies of various translated versions have also mostly reported acceptable validity coefficients. In India Shamasundar et al. (1986, cited in Gureje and Obikoya, 1990) reported a sensitivity of 87% and specificity of 93% for the 12-item version of the GHQ. Vazquez-Barquero et al. (1985) refer to two Spanish studies which obtained sensitivity scores of 81% and 77.9% and specificity scores of 88.2% and 81.4%. Aderibigbe and Gureje (1992) using the Yoruba version in Nigeria found sensitivity of 82% and specificity of 85%. By comparison, Gureje and Obikoya (1990) obtained lower coefficients with the Yoruba version of the 12-item GHQ in Nigeria, with sensitivity and specificity at 68% and 70% respectively.

3.7 South African studies that have employed the GHQ

There are at least five studies that have employed the GHQ with South African samples. A brief description of each of these studies is given here with the intention of examining the aspects listed below:

1. Area of investigation
2. Type of sample used
3. Version of GHQ used
4. Measures taken to adapt the GHQ for local setting
5. Method of administration
6. Method of scoring
7. Performance and utility of the GHQ

A review of this nature is deemed necessary in order to make a critical assessment of the South African studies and to draw observations as outlined in the subsequent section.

O'Neil (1988) used the GHQ to examine the mental health status of South African shop stewards. She administered the 30-item version of the GHQ to a sample of black shop stewards. The GHQ was translated into Zulu by two bilingual research assistants, using a single back-translation procedure. The Zulu version was piloted on a sample of five Zulu speaking worker representatives to ensure that it was comprehensible. The author reports that a Yes/No response format was employed, because of a lack of familiarity with the Likert scaling concept amongst the sample.

Leeb (1986) employed the GHQ to evaluate the psychological impact of unemployment. She administered a Zulu version of the 30-item GHQ to 126 people in the black peri-urban, working class community of Mphophomeni, Natal. The author states that the GHQ was chosen because it "has been proven to be effective over most populations and not to differ significantly when affected by variables such as race, age or sex. Other scales were considered, but the GHQ appeared to be most suited" (p4). The items were

presented verbally and responses were recorded by an interviewer. The scoring system was the simplified Yes/No response format for all items. Leeb (1986) discusses seven items that were found to be problematic in the study and suggests refinement of the questionnaire. The GHQ was nevertheless able to differentiate between the stress levels of employed and unemployed groups.

Turton, Straker and Moosa (1991) used the GHQ to investigate the experiences of violence of two cohorts of township youth. The shortest form of the GHQ, the 12-item version was administered to two groups of standard 10 pupils at a secondary school in the black township of Alexandra in the Gauteng province. It would seem that the questionnaire was administered in its original English form. The questionnaire was completed by the respondent, in the presence of the researcher. Each item of the GHQ was coded as either 0 (negative) and 1 (positive). The researchers used a cut-off score of three or more positively coded items, as an indication of the presence of distress. The study found that the results of the GHQ concurred with those of an independent symptoms checklist, and was also related to other responses indicating stress-related affects, anger and dysphoria.

Pillay, Magwaza and Petersen (1992) used the GHQ to examine the psychological sequelae of civil conflict. They administered the GHQ-30 to a sample of primary caregivers in the African township of Mpumulanga, in KwaZulu Natal. The report of the study gives no details regarding the version used, adaptation measures undertaken, if any, or of scoring methods employed. Personal correspondence with the first author, revealed that a Zulu translation of the GHQ was used by interviewers who verbally presented items and recorded the responses of the respondents. The authors report a significant correlation between scores on the GHQ and an independent measure of Post Traumatic Stress Disorder.

Moodley (undated) employed the GHQ to assess the impact of forced removals. The GHQ was administered to a sample of rural African

people in KwaZulu Natal. The researcher did not use any of the
predefined versions of the GHQ, but chose a selection of items
from the GHQ-60. Due to methodological problems the study remains
incomplete.

3.8 Observations from South African studies using the GHQ

All the studies reviewed above, administered the GHQ to African
samples. It is most unlikely that English was a second language
for these respondents. This was certainly the case with the Leeb
(1986), Pillay et al. (1992), and Moodley (undated) studies. All
of the studies except that of Turton et al. (1991), were
conducted in Natal and involved Zulu speaking samples. None of
the studies employed the full 60-item version of the GHQ, the
most popular version being the 30-item. This is probably because
the GHQ was part of a larger battery of instruments (for example,
O'Neil, 1988; Turton, et al., 1991; Pillay et al., 1992) or was
attached to a long questionnaire (for example, Leeb, 1986).

The four studies that used some form of a Zulu version of the
GHQ, adopted a simplified response format of Yes/No (Yebo/Cha, in
Zulu) rather than the more complex Likert range of responses in
the original English version. The Zulu-sample studies also show a
deviation with regard to administration. They all involve a
verbal presentation of items. This was dictated by conditions of
high levels of illiteracy among the samples, a problem
experienced in other African sites, such as Nigeria (Aderibigbe
and Gureje, 1992), where self-administration was impossible.

A notable feature of the SA studies, is the lack of attention
given to the issue of the appropriateness of employing the GHQ
locally. Perhaps, the most blatant disregard for cross-cultural
applicability is that demonstrated by Turton et al. (1991), who
employed the original English version of the questionnaire, with
black secondary school pupils. Of the studies that employed
translated versions, very little has been done to check the

fidelity of the translations. O'Neil (1988) reports one step of back-translation and a small pilot. While she attempts more than any of the other studies, in terms of ensuring some measure of equivalence, the steps taken are not exhaustive of the available methods. Results are therefore still subject to translation error. Back-translation on its own is not a guarantee of high fidelity. (see section on translation). The Leeb (1986) study makes no reference to translation measures, nor is any pretesting evident. Leeb (1986) does however, examine the properties of individual items after having applied the GHQ.

Earlier discussion has shown that the GHQ is a widely used instrument internationally. Given the scarcity of available instruments, its use in this country is likely to continue. It is evident from the review of the studies above and from many other studies that employ foreign instruments, that concerns about local calibration and establishing psychometric equivalence, are not high on the agenda of researchers. This is often because the instruments used in studies are merely a means to some more important or interesting end. The studies reviewed here are all valuable in that they examine areas of important social concern and in many ways, redress the past imbalances of research in this country.

However, despite noble concerns and a display of credible responsiveness to the challenges of problems in South Africa, such research is compromised by the inadequate attention given to the use of instruments of foreign origin. In the interest of gaining greater confidence in our results and in order to eliminate the possibility that findings are an artifact of an inappropriate instrument, there is a need for preliminary research on the instrument itself. It is suggested that such instruments be subjected to an extensive research programme. Such a programme for the GHQ has been discussed (p14). Note that the present study is seen as a contribution to the initial stages of such a research programme.

Chapter 4

Theoretical issues in cross-cultural psychology

Cross-cultural psychology is "concerned with the systematic study of behaviour and experience as it occurs in different cultures, is influenced by culture, or results in changes in existing cultures" (Triandis, 1980, p1). Research in cross-cultural psychology often involves a comparison of two or more cultures with regard to some aspect of psychological functioning or behaviour. At other times it gets its definition because the researcher and the research participants belong to different cultures, and where the researcher is usually interested in examining constructs or behaviour that are known to exist in his or her own culture (see Irvine and Carroll, 1980). A further case for the term cross-cultural, is when the instruments used in a study, were developed for a culture that is different to the one under examination (see Sang, 1992).

The present research is envisaged as belonging to this area of psychology because the instrument under examination, the GHQ, was developed for and on a culture that is very different to the one it is intended for use with in this study. Apart from the instrument being foreign, the psychological constructs associated with it, may themselves be foreign to the target culture. Triandis (1980) explains that cross-cultural psychology is defined by its methodology rather than by its theory. It is argued that the methods developed for cross-cultural research and the attendant special requirements imposed on the researcher, are both imperative and valuable to the present study.

Culture is a difficult concept to define. Poortinga (1992) cites a study by Soudijn et al. which analysed 128 definitions of culture and concluded that there was no one definition which could be considered best.

A classic definition of culture is: 'that complex whole which includes knowledge, belief, art, morals, law, custom, and any other capabilities and habits acquired by man as a member of society' (Tyler, 1874, cited in Rack, 1982, p13).

For the purpose of the present study, a more pertinent definition, given the specific reference to the important role of language in culture, is:

'the categories, plans and rules people use to interpret their world and act purposefully within it ... the grammar used to construct and interpret behaviour. Culture is learned as children grow up in society... Culture is a plan for behaviour, not behaviour itself.' (Spradley and McCurdy, 1974, cited in Rack, 1982, p13)

A major purpose of cross-cultural psychology is to examine the generality of psychological laws (Triandis, 1982). Much of recent research has therefore been occupied with the search for universals or what is known as 'etic' aspects. Owen (1991) notes that there is more similarity between the cognitive structures of different cultural groups than is generally believed. Rack (1982) states that studies have usually shown that depressive illness exists wherever it is looked for, if the questions are framed in the right manner. However, although it is evident that mental illness exists throughout the world, its manifestations are not the same everywhere. Furthermore, different cultures and eras set different boundaries on what is to be regarded as mental illness and what should be done about it (Rack, 1982).

Utilising methods developed in cross-cultural psychology is not in itself a safeguard to good research. The researcher has to be cautious of ethnocentric bias at all steps in the study. One example of ethnocentricism would be, when the researcher, armed with foreign tools and constructs applies his skills to a local group of people and makes inferences using his/her own culture as a term of reference. Schweitzer, (1977) examines a range of early

South African studies which examined 'hallucinations' experienced by Black patients. Apart from offering some memorable quotes from these early ethnocentric studies, he warns that while the studies may be considered to be methodologically sound, they failed to relate the experiences described as being meaningful within the individual's own cosmology.

X The perspective of cultural-relativism on the other hand involves a deep understanding and respect for the culture under study and an acknowledgement by researchers, that their own values are culture-bound. Rack (1982) advises that "if we start, as we all do, from a position of ethnocentrism, a large dose of cultural relativism is a useful purgative ..." To know a culture we must recognise its local, transient, and unique characteristics. And it is only in relation to this culture that we are entitled to say whether or not behaviour is abnormal" (pp 9-15). Schweitzer (1977) observes that cultural relativists acknowledge that there are no universal norms which can be applied to all societies. The present study accepts the position of the cultural-relativist view and reflects this in the 'research programme' advocated for the GHQ (p14).

4.1 Language and Linguistic factors

The Spradley and McCurdy definition of culture cited above, illuminates the important role of language in any culture. There are two conceptions of the role of language in the study of theoretical constructs. The extreme Whorfian position states that individuals who speak different languages live in separate worlds. In this view, language is seen as the filter between man and the world. The strong linguistic position states that very high fidelity translations from a source language to a target language provide a sufficient basis for cross-language and cross-cultural assessments and comparisons. The latter position subscribes to the view discussed earlier, that there is a broad degree of the unity of mankind (Hulin and Mayer, 1986).

Lonner (1980) notes that, "cross-cultural psychologists are generally not interested in the origins of language or the discipline of linguistics, they are more interested in the affective meanings that words carry in specific social contexts, and the problems associated with translation so that linguistic stimuli can be made as equivalent as possible" (p160). If this is the case, then a large part of the problem that exists, is the different vocabularies that exist in different cultures.

Rack (1982) notes that each culture develops a rich vocabulary around the issues which seem particularly important to that culture at that time. Nel (1994) observes that language is the most important single moderator of test performance. In examining the terminology used to describe depression, Rack (1982) shows how rich the English language is, in describing mood states. Although not exact synonyms, alternatives can be despondent, despairing, disconsolate, dispirited, disillusioned, gloomy, melancholy, miserable, morbid, morose, unhappy, sad, and so on. He notes that no such vocabulary exists in non-European languages and offers the Yoruba example where one word suffices for both 'angry' and 'sad'. The crucial question is however, is the experience the same even if the vocabulary is different.

The presence of large numbers of metaphors in a vocabulary and differences between systems of metaphorical representation pose a particularly difficult methodological problem for cross-cultural research. Dunnigan, McNall and Mortimer (1993) who looked at translation problems posed by metaphorical nonequivalence between two cultures under study, notes that "the question remains whether multilingual research that relates to highly metaphorized cultural domains can produce comparable data for culturally different groups" (358). They believe that the answer, can only be decided on a case-by-case, or even variable-by-variable basis determined by careful semantic analysis. The authors provide a very useful example of such a semantic analysis.

There are some items in the original English version of the GHQ that consist of metaphors. Special attention will need to be paid to such items in translation and assessment of equivalence. Dunnigan et al. (1993) state that "Researchers may discover that contrasting metaphors in different languages have essentially the same referents, ... Despite this referential similarity, direct translation from one language to another will not always produce interview instruments that are semantically matched item for item, and a separate inventory of culturally appropriate questions may have to be developed for each subject group" (p359).

There is particular relevance and wisdom in the following old quotation; "If a man does not keep pace with his companions perhaps it is because he hears a different drummer. Let him step to the music he hears, however distant" (Thoreau, 1817-62, cited in Rack, 1982, p100). One implication of this is that psychology must take care not to diagnose someone as mentally ill simply because the person is out of step with society. Another, is that the profession must guard against ethnocentric bias in imposing a system of understanding and its attendant tools on other cultures.

4.2 Responses from bilinguals

As discussed later in the methodology section, the final step for assessing equivalency of the Zulu version of the GHQ, involved administering both language versions of the questionnaire to Zulu-English bilinguals. This practice is commonplace in cross-cultural research and is recommended as an important translation step in the works of Brislin (1970) and Werner and Campbell (1970). Bilingual samples are often employed in studies that have assessed translations of the GHQ (Chan, 1985; Sriram et al., 1989).

Although the practice of using bilinguals samples is widespread, Church, Katigbak and Castaneda (1988) note that little is known about the effects of collecting data in a native (primary) versus a second language. The suggestion that language can affect the responses of research participants, is based on studies which have found differences in the responses given by bilinguals to identical items when asked in the different languages they speak (Yang and Bond, 1980; Bond and Yang, 1982; Church et al., 1988; Tyson, Doctor and Mentis, 1988). While such discrepant responses have been obtained from bilingual respondents, there is no consensus explanation of this phenomenon and little replication of existing findings is available. Explanations accounting for these findings have been expressed in terms of social reference groups, language acquisition and semantic constructions of languages.

4.2.1 Social reference groups explanations

Among the social psychological explanations espoused to explain these differences, are the 'cross-cultural accommodation' hypothesis, the 'ethnic affirmation' hypothesis and the 'social desirability' hypothesis. The one most frequently espoused is the accommodation hypothesis which states that when individuals acquire a language, they also acquire attitudes, values and role expectations associated with that language. Learning a language therefore involves a process of acculturation, the result of which becomes evident when a bilingual offers a response that is appropriate to the particular culture in whose language the question is asked.

According to the alternative 'ethnic affirmation' hypothesis as proposed by Yang and Bond (1980), a bilingual's ethnicity becomes salient when he or she is asked to respond in their second language, causing the individual to endorse their home culture values to a greater extent. Yang and Bond (1980) compared the responses of Chinese students to a questionnaire presented in

either Chinese or English, and found that Chinese bilinguals responded in a more traditionally "Chinese" direction when given the questionnaire in English than when given the equivalent Chinese version of the questionnaire. They argue that the use of a second language (English) cued their Chinese respondents to their ethnicity.

In a later study, Bond and Yang (1982) report evidence of both affirmation and accommodation. Their explanation of the later finding is based on an 'importance' factor. They found that the greater the importance of the value to the respondent's culture, the more likely it was that ethnic affirmation would be displayed. In the study the importance of values was independently established. They therefore concluded that when an individual is asked to respond about a value or belief that is perceived to be important to the individual's culture, ethnic affirmation is most likely to occur, whereas when the questions are of a less important nature, cross-cultural accommodation is more likely to occur.

While both the studies by Bond and Yang have shown evidence of this type of ethnic affirmation, Tyson et al. (1988) note that no other study has replicated the effect. The lack of replication of language effects is also bemoaned by Church et al. (1988) who state that language "differences have often been unsystematic and interpreted post hoc" (p179). Furthermore, Yang and Bond (1980) report that ethnic affirmation occurs when the respondent is answering in their second language. Marin, Betancourt and Kashima (1983) report evidence of ethnic affirmation, but only when their respondents were answering in their home language. It has been suggested that the results obtained by Bond and Yang, were a function of their methodology (Tyson et al., 1988).

A third explanation for the discrepant responses of bilinguals when asked the same question in different responses, is explained by the social desirability hypothesis. Research has shown that when respondents are asked to complete an instrument in their

second or non-native language, they present themselves in a more
socially desirable manner. Marin et al. (1983) discuss a study
conducted with Greek bilinguals in which low correlations were
found between the Greek and English answers given by the same
respondents to those items that differed in social desirability
in Greece and the United States. The Greek bilinguals presented
the most socially desirable responses when answering in their
second language, rather than in their mother tongue.

Based on a study conducted with English- and Afrikaans-speaking
South Africans, Tyson et al. (1988) add an important proviso to
the social desirability hypothesis advanced by Marin et al. They
argue that social desirability effects are only likely to occur
when a status difference exists between the two cultures under
study. Tyson et al. note that in South Africa, English and
Afrikaans enjoy relatively equal status in both official and
unofficial terms and may therefore have no instrumental value
attached to either. This is offered as the reason for the absence
of social desirability effects in the Tyson et al. study, because
unlike in other studies, English may not be perceived to be a
prestigious language that is associated with social rewards.
Given the legacy of apartheid and the existing racial disparities
discussed earlier (p2-7), it is unlikely that the same
circumstance of relative equality would exist between English and
other African languages, for example Zulu.

4.2.2 Language acquisition explanation

Tyson et al. (1988) state that the kind of discrepancies between
a bilingual's responses in his or her primary language and
another language "are the result of the individual engaging,
albeit unconsciously, in impression management" (p415). They are
therefore social in nature. An alternative to the social
psychological explanations described above, states that the
discrepancies are a function of language acquisition. According
to Ervin and Osgood (1954, cited in Tyson et al., 1988),

bilinguals can be differentiated according to the manner in which they acquired their two languages. They distinguish between 'compound' bilinguals and 'coordinate' bilinguals. Compound bilinguals acquire their languages in settings in which both languages were used interchangeably, hence a fusion of the two meaning systems occurs because both the languages were learned in the same setting. Coordinate bilinguals acquire their languages in separate contexts, which are populated by speakers of only one or the other of the two languages. The separate contexts therefore create two distinct meaning systems.

Tyson et al. (1988) state that the implication of the compound-coordinate categorisation system "would be that responses of compound bilinguals to the same questionnaire given in either of their two languages should be the same, whereas responses of coordinate bilinguals may differ" (p415). Most, if not all of the bilingual respondents in the present study could be considered to be coordinate bilinguals. It is safe to assume that Zulu would have been the language learnt first and spoken predominantly at home, and English the language acquired some years later with the start of formal schooling.

4.2.3 Semantic construction

A further explanation of the non-social type, states that language effects could also emerge from differences in the semantic makeup of the individual languages. It is argued that differences in the prevalence and saliency of concepts could result in language effects (Church et al., 1988).

The lack of consensus regarding the effect of language on the responses of bilinguals is perhaps best summarised by the following two quotations;

"Our results suggest a `cup half full, half empty' phenomenon regarding the extent to which language of data collection affects the results of psychological research" (Church, et al., 1988)

"The results do not support the statement that `at this point it seems certain that bilinguals differ in their responses when they answer a questionnaire in their two languages' " (Tyson et al., 1988)

It would appear that the jury is still out regarding the discrepancies in bilinguals' responses reported in some studies. A point worth noting however, is that the studies discussed above all tended to investigate the issue via the use of questionnaires that tapped values, morals and personality aspects. Such items could be argued to be more culture-bound than items in a health measure such as the GHQ. Since language is central to the culture in which it is spoken, it is therefore more likely that such studies would evoke culturally-mediated responses. It would be interesting to attempt to replicate the language effects reported, via the use of questionnaires that tap less culture-bound aspects of life.

Chapter 5

Theoretical models of measurement equivalence

The use of tests in cultures other than those for which the test was developed, introduces serious considerations for the researcher. Paramount to these considerations is rigorous translation (discussed in detail in Chapter 6) and thereafter providing evidence of measurement or psychometric equivalence. The failure of many studies to do this was expressed earlier (see p9). The requirement of psychometric equivalence is usually demonstrated via item analyses and by the translated test's indices of reliability and validity in the new setting. Research has also pointed to the requirement of assessing culturally transported tests to identify items that do not function equivalently for both cultural groups (Ellis, Becker and Kimmel, 1993). As defined by Drasgow (1987, cited in Ellis, et al., 1993), a transported test can be said to fulfil the requirement of measurement equivalence "when individuals with equal standing on the trait measured by the test but sampled from different sub-populations have equal expected observed test scores" (p133).

5.1. Classical theory

The majority of past studies that have demonstrated indices of equivalence, have done so within the framework of classical measurement theory. Hambleton, Swaminathan and Rodgers (1991) note that classical measurement theory and its attendant procedures for constructing and evaluating psychological tests and, interpreting test scores, have for a long time, served the testing branch of psychology well.

Common statistical procedures applied within classical theory, include examinations of mean scores and standard deviations, factor analysis and, an array of correlation analyses. Correlation analyses have focused both on total scores and on

items scores. Hulin and Mayer (1986, p84) list the four most commonly used procedures for detecting measurement. They are;

a) item extremity or p value
b) the point-biserial correlation between item response and total test score;
c) mean of examinees' scores;
d) standard deviations of examinees' scores; and
e) factor analyses of item response covariances

Hulin and Mayer (1986) note that the problem with many of these procedures, is their subpopulation dependence. They therefore recommend these classical statistics be used when making within group comparisons, but may not be appropriate for comparison across groups. As noted earlier, one of the aims of this study to initiate the process for establishing an equated Zulu version of the GHQ which would find its main value, in South African studies that fall in the area of within-group studies. The argument is, that so little is known about mental health within the Zulu-speaking population, that cross-cultural studies would be superfluous and a luxury that would not serve development in the country at present.

5.2. Item Response Theory

This mathematical approach has been variously discussed in the literature as item response theory (IRT), latent trait theory, and item characteristic curve (ICC) theory (Anastasi, 1982). The approach is being heralded as a revolution in educational and psychological measurement, and seen as a significant improvement on classical measurement theory (Hambleton et al., 1991).

In advancing support for IRT and noting the shortcomings of the classical model, Hambleton et al. (1991, p.5) point to the following characteristics of the IRT approach:

a) item characteristics are not group dependent
b) scores describing examinee proficiency that are not test-dependent
c) the model is expressed at the item level rather than at the test level.
d) the model does not require strictly parallel tests for assessing reliability, and
e) the model provides a measure of precision for each ability score.

A fundamental feature of the IRT approach is that item performance is related to the estimated amount of the respondent's "latent trait", where "latent trait", refers to a statistical construct, not a psychological entity (Anastasi, 1982). Item characteristics are plotted from mathematically derived functions, rather than more directly from the empirical data used in item-test regression curves.

Different IRT models have been developed, each using different mathematical functions, based on diverse sets of assumptions (Hambleton et al., 1991; Anastasi, 1982). A factor analytic model is one type of item-response model, which describes the linear relationship between item responses and psychological constructs (Parsons and Hulin, 1982). Increasing use is being made of ogive or logistic models.

IRT procedures are based on the assumption that a unidimensional trait is being assessed by the instrument under examination (Hambleton et al., 1991; Parsons and Hulin, 1982). No factor analytic studies have been conducted with the GHQ in Zulu-speaking populations. Since the number of factors that the GHQ may be assessing in the present population is therefore unknown, the assumption of unidimensionality cannot be easily made.

Although the GHQ was originally developed with four sub-scales (Goldberg, 1972), a number of studies have pointed to an acceptable single factor of general well being (see section on factor analysis p.34).

One area of testing in which IRT is receiving considerable support, is for the purpose of checking translation fidelity of instruments. An aspect of IRT referred to as Differential Item Functioning (DIF), 'item bias' in classical terms, is used to determine the quality of translations. Items that have been inaccurately translated are likely to show DIF. Apart from translation error, translated items displaying DIF, may also arise from (a) differences in cultural relevance or meaning of the item, or (b) differences in culturally specific knowledge Candell and Hulin, (1986).

In a critique of item response theory, Anastasi (1982) states that there is need for much more checking of mathematically derived values on live data, rather than with artificial data and computer simulation as has been the case thus-far. A research programme with the Job Descriptive Index (Hulin and Mayer, 1986; Candell and Hulin, 1986 and Parsons and Hulin, 1982) appear to have taken up the challenge. Anastasi (1982) also notes that psychometricians question the very applicability of IRT procedures such as ICC techniques, to psychological tests on theoretical grounds. She offers the example of the central assumption of unidimensionality, which cannot be demonstrated by the usual factor-analytic procedures, because item intercorrelations are likely to be curvilinear. Of greater possible relevance to the present study, Anastasi (1982) warns that "the same items may involve a different mix of abilities when performed by persons with different experiential backgrounds or by the same person at different stages of learning" (p.215).

5.3. Theoretical models applicable to the present study

A overwhelming majority of studies that have attempted to demonstrate equivalence of different forms of the GHQ, have employed procedures of the classical model. A notable exception is that of Mari and Williams (1985), who compared the validity of the GHQ-12 and the Self-Reporting Questionnaire (SRQ), using Relative Operating Characteristics (ROC) analysis.

Candell and Hulin (1986) employed IRT procedures for assessing the translation of the Job Descriptive Index (JDI) into Canadian French. They interpret the finding of just a single item with DIF due to translation error, as a testimony to the effectiveness of the translation/back translation process. In a similar study in which the JDI was translated into Hebrew, Hulin and Mayer (1986) found 21 of the 66 items to have significant bias. The present study can be envisaged as paying focal attention to the important preliminary step to a possible IRT analysis, that is, developing a translated version of the GHQ of high-fidelity (see methodology).

For the present study, a reliance on bilingual samples is made. DIF analysis in IRT routinely compares two groups that are usually of primary interest ie., the focal and reference groups Ellis and Kimmel, (1992). An IRT analysis will require the two versions of the GHQ (original and translated) to be administered to separate English-speaking and Zulu-speaking samples. Furthermore, the English-speaking sample or reference group would be used as the standard against which the Zulu-speaking focal group is compared. However, the psychometric status of the GHQ in both focal (Zulu) and reference (English) versions is unknown in the present context.

IRT analyses will also necessitate the use of samples much larger than that involved in the present study. Candell and Hulin (1986) report in their study, that confidence in the parameter estimates (IRT) and chi-square could be increased by using samples larger

than those that they employed. Their study used groups of 295 and 213 participants. The samples (1632 and 308) employed by Hulin and Mayer (1986) appear satisfactorily larger. The IRT procedure of DIF analysis requires that the items be dichotomously scored. The GHQ has both a Likert and binary scoring method. Although the likert method has been shown to offer a better spread of scores for psychometric analysis, the binary method would most likely have to be employed for IRT analysis at present. The value of multi-point IRT programs that are being developed would seem to require more assessment with genuine data.

A further consideration is presented by Anastasi (1982), in that IRT involves "a class of mathematically sophisticated procedures", with "extensive computations", that have only become practicable with the availability of high-speed programs (p91). Hambleton, et al. (1991) describe a number of computer programs that have been specifically developed for IRT analysis.

In chapter one, a long-term research programme has been proposed for the GHQ and indeed, for all other foreign instruments intended for use in South Africa. Along with the establishment of local norms for the GHQ and a range of studies assessing its validity, the more ongoing process of developing a psychometrically equivalent Zulu version of the GHQ, would be well served by the use of IRT procedures at various appropriate stages in such a programme. This general area of research seems likely to be a worth-while future use of IRT methods.

Chapter 6

Translation theory and methods

This section offers an examination of the theory on translation and sketches an outline of the main methodological models available for research involving translation.

Over four decades ago, Richards (1953, cited in Brislin, 1980) asserted that translation is "probably the most complex type of event yet produced in the evolution of the cosmos (p.25)." The hyperbole conveys a very vivid characterisation of the intricacy involved in translation work in general and for psychological assessment purposes in particular. More recently, studies still report agreement with the essence of Richards' view regarding the complexity of translation (Hulin and Mayer, 1986, Candell and Hulin, 1986). Lonner (1990) states that many hours of careful and dedicated research may be needed to make even a brief questionnaire appropriate for culture-comparative research. Sartorius and Kuyken (1994) offer the criticism that the translation of health status measures, has to date, failed to acknowledge the complexity of the translation process. They identify the following reasons for the complexity;

"Differences exist between cultures in the construction of health and illness, levels of literacy, reading level, concordance between written and spoken versions of the language, taboo subjects, and social desirability effects. Furthermore, certain features of language, such as idiom, are very difficult to translate, and abound in some health status instruments." (p.4)

All of the points made by Sartorius and Kuyken (1994) above, are relevant to the present study in that they refer to some of the key criteria by which one may judge the similarities or differences between the GHQ's country of origin (United Kingdom) and the site of intended use of a Zulu version, namely, South Africa.

Despite the complexity involved in the translation process, significant methodology gains have been made, particularly in the field of cross-cultural psychology, to allow researchers to proceed with translating instruments. The early works of Brislin (1970) and Werner and Campbell (1970) appear authoritative on the subject of translation of psychological materials. Other useful contributions have been made by Sechrest, Fay and Zaidi (1972), Sartorius and Kuyken (1994) and Retief (1988). Brislin (1980) outlines Casagrande's four "ends" or types of translation, namely "pragmatic, aesthetic-poetic, ethnographic and linguistic", as a useful introduction to the topic of translation;

6.1 Casagrande's ends of translation

Pragmatic translation, is the sort utilised by repair manuals and similar documents, where accuracy of the information that was meant to be conveyed, is the primary interest to the translator. Unlike the remaining three types of translation described below, with pragmatic translation, the translator is not interested in aspects of the source language version, but only that the information is conveyed in the target language.

With aesthetic-poetic translation on the other hand, the translator focuses not only on the information in the message, but also on the affective and emotional shades expressed and the feeling of the original language version. Translations of literature would be an example of aesthetic-poetic translation. Werner and Campbell (1970) call this "asymmetrical" translation because loyalty to the source language dominates.

Within the third type, ethnographic translation, an attempt is made to delineate and explain the cultural context of both the source and second language versions. Here translators pay specific attention to the way in which words are used in an attempt to contextualise the words to the cultures which use the source and target languages. Brislin (1980) cites a description

of a particular translation of the Bible as a good example of
ethnographic translation. Mardell-Czudnowski et al. (1986),
note that before the DIAL-R was normed in Taiwan, it had to be
revised to meet the cultural and social background of this area.
They report satisfactory results when the test was piloted on a
small group of Chinese children, following the revision and back-
translation processes.

Linguistic translation, focuses on the structure of language. In
addition to grammatical form, attention is paid to "equivalent
meanings of the constituent morphemes of the second language"
(Casagrande, cited in Brislin, 1980, p428).

While noting that a single translation exercise may be
categorised into more than one of Casagrande's four types,
Brislin (1980) suggests that an awareness of the different types
of translation should assist the translator in prioritising and
choosing specific goals for a particular translation exercise.
The present attempt at translating the GHQ into Zulu could be
seen to be governed by the objectives of at least two of
Casagrade's four types. A central objective is that of attainment
of equivalent meaning, namely, that the information contained in
the source language is conveyed in the second language (pragmatic
translation). Beyond this, attention would be given to ensuring
that the words chosen for the translated version, are harmonious
with the culture that speak that language (ethnographic
translation).

While writers in the field of psychology such as Brislin (1980)
and Retief (1988) find Casagrande's "ends of translation" a
useful introduction to topic of translation, Casagrande system of
categorisation as evident in the discussion above, has much
broader application. In essence it encompasses translation in all
its forms and across a wide range of disciplines. A more focussed
discussion of translation, looking specifically at the goals of
cross-cultural psychology is offered by Sartorius and Kuyken
(1994). Their "four approaches to translation" is examined here,

not only because it is more focused and relevant to the present project, but also to avoid possible confusion regarding some of the terminology used, such as the "pragmatic approach" of Sartorius and Kuyken (1994), which is altogether different from the "pragmatic translation" discussed by Casagrande.

6.2 Sartorius and Kuyken's approaches to translation

The "ethnocentric approach" describes the practice where a foreign instrument is translated into the language of a second culture, without prior assessment of the appropriateness of the instrument for the second culture. This approach which describes the conduct of a number of researchers who employ foreign instruments, makes the assumption that the conceptual dimensions of health that underpin the source instrument, are equally applicable to the target culture. Ward and Sethi (1986) complain that the "most outstanding flaw in the majority of cross-cultural BSRI (Bem Sex Role Inventory) research is the tendency for investigators to simply administer the personality inventory and compare results with original American data without examining the appropriateness and relevance of item selection for the host culture" (p302). The "ethnocentric approach" also assumes that the source and target languages are sufficiently similar to allow for meaningful translations and that the method of questioning involved in the source instrument, is acceptable in the target culture. (Sartorius and Kuyken, (1994)).

Sartorius and Kuyken, (1994) warn that the "principal danger of the ethnocentric approach is the distortion in results that results from the use of inappropriate health constructs, which are valid in the source setting but not in the target setting." (p5).

The "pragmatic approach" involves looking for common ground between the source and target cultures, and measuring and translating only those health constructs that both cultures

share. The decision of whether the two cultures share a semantic space regarding the health construct is based on the advice of key informants such as bilingual health professionals and lay workers (Sartorius and Kuyken, (1994)).

The "emic plus etic approach" involves finding and translating concepts that are common to both cultures (etic aspects) and then looking for culture-specific aspects (emic) of the target culture, with the view of measuring them. Brislin (1986) offers a detailed description of this approach as employed by Miller et al., in a cross-cultural study of authoritarianism-conservatism. The study used an existing instrument which allowed for comparisons with previous studies, but the researchers also modified items and added new items, thereby also allowing for culture-specific enquiry. The approach has definite value in terms of capturing culture-specific aspects of the target culture; however one is in essence left with a new instrument and comparisons with results of studies using the original instrument are not straightforward.

The "re-presentation approach" involves an interpretation of the source language items which is guided by the translator's understanding of the target culture. The approach shares a similar objective with the decentering process described by Werner and Campbell (see p68), with the main difference being that the re-presentation approach accepts the likelihood that the process would result in two similar instruments whose application will not present comparable data.

6.3 The quest for equivalence in translation

The goal of attaining equivalence between the source and target language versions of an instrument is central to any translation effort. Equivalence as described in the translation literature is however, a multi-dimensional concept. There are also no widely-accepted criteria for translation equivalence against which a

given study may be gauged (Brislin, 1970). A number of writers have developed lists of the different types of equivalence. Although the terminology varies, there is some overlap in the actual concepts being referred to.

6.4 Types of equivalence in translation

Berry and Dasen (cited in Retief, 1988) suggested three kinds of equivalence that could be demonstrated in a translation exercise:

1. Functional equivalence, which is shown to exist when two or more behaviours in two or more cultural systems, relate to functionally similar problems.

2. Conceptual equivalence is the requirement governing the meaning of research materials or behaviour which must be equivalent before comparison between cultures is possible.

3. Metric equivalence refers to the psychometric properties of two or more sets of data from two or more cultural groups that exhibit essentially the same coherence or structure.

Sartorius and Kuyken (1994) note that the aim of translating health status measures, is to maintain conceptual, semantic and technical equivalence between both language versions of the instrument. Here, conceptual equivalence refers to the same concepts underpinning individual question items in both language versions of an instrument; semantic equivalence refers to a common set of connotations and denotations of words used in both language versions; and technical equivalence refers to both equivalence of technical aspects of language and the appropriateness of the nature and mode of questioning required by the instrument in both cultures.

Sechrest et al. (1972) and Retief (1988) provide similar lists of types of 'equivalence' that serve as a useful practical guide for translators. The list formed the basis for instructions that were be given to the translation team in the present study and is therefore outlined in Appendix A.

6.5 Translation problems

The criticism that inadequate attention has been given to translation, has already been submitted. However, even when attention is paid to the problems surrounding translation of research instruments, this attention too has been criticised for having too narrow a focus. Researchers often tend to address themselves only to the problems of phrasing questions contained in the instrument. This sometimes becomes the sole focus of their translation efforts which often involves establishing equivalency via various statistical techniques.

Sechrest et al. (1972) comment on three other problem areas in translation, all of which do not receive much attention from researchers. They firstly note that most research requires some orientation to the study be given. These introductory statements, whether in oral or written form, usually presents a rationale for the study and pledges of confidentiality et cetera, and should necessitate precision in translation. Sechrest et al. (1972) note that there are "no instances known to us in which an investigator specifically mentions such a problem, let alone a solution to it" (p42).

The second type of problem relates to the translation of instructions specific to different types of tasks or measures involved in an instrument. Sechrest et al. (1972) note that here too, very few investigators pay much attention to this aspect of the instrument. They state that "very few investigators can be described as having been sufficiently wary to make us totally confident in their findings" (p42). They cite cross-cultural

research involving the Rod-and-Frame in which two of the most difficult problems encountered in translating instructions, were finding a Tagalog (Philippine language) word for "upright" and an Urdu word for "rod". In this regard, it is important to note that the instructions for the GHQ are both short and simple. These instructions will however be subjected to the same amount of rigour and scrutiny in translation, as will the actual questions in the test.

The final type of translation problem which is often inadequately addressed, involves translation of responses. This is a particular problem for open-ended questions and interviews. Fortunately this type of problem would not affect the GHQ. Completing the GHQ does not require any written responses. Respondents need to only tick the appropriate responses which are all provided on an answer sheet.

6.6 Translation procedures

A number of procedures are available for producing translations of high quality and equivalency. Despite the availability of such procedures, Brislin (1970) notes a failure amongst researchers to make use of them. In support of this view, he cites a study by Campbell et al., who reviewed 80 articles in the Journal of Social Psychology and Public Opinion Quarterly, both considered to be major publication outlets for cross-cultural research, and found that 61 articles gave little or no information on the translation procedures that were employed. The lack of information was such that poor translations could not be excluded as a source of data contamination. While this latter set of studies are dates, it has previously been shown that a similar problem is evident with South African studies in general (see p.9) and with South African studies that have employed translated versions of the GHQ (see p.40).

Some of the translation techniques discussed in the works of Brislin (1970, 1980, 1986), Werner and Campbell (1970) and Sechrest et al. (1972) are discussed here. Most authors recommend the use of a combination of the various techniques in any translation effort. The present study will employ a number of the recommended techniques.

6.6.1 Back-translation

As noted earlier, a number of studies have employed the back-translation procedure. The procedure should been seen as a necessary, but not sufficient step.

In this procedure, two bilinguals who are both competent with their two languages are given the task of translating the instrument. The first bilingual translates from the source to the target language. The second bilingual having not seen the original source version, translates the target version produced by the first bilingual, back to the source. The procedure leaves the investigator with two source versions, which he or she may compare. If from the comparison the two source versions are considered to be identical, then the target version from the middle of the procedure is deemed to be equivalent to the original source version.

Research reporting successful use of back-translation

Werner and Campbell (1970) report successful use of the back-translation method with Navajo respondents translating simple English passages. Brislin (1970) also cites early studies by Fink and by Sinaiko who made use of the back-translation technique. A number of studies involving the translated versions of the GHQ (see p.36), also reported successful use of back-translation.

Of the available techniques for translation, back-translation is most commonly used. Some of the main reasons for its popularity are that:

a) back-translation allows for some degree of quality check and is therefore more superior to a simple translation from source to target language.

b) the researcher is often not familiar with the target language and the culture of the people who speak it. Back-translation allows the researcher to be involved in assessing the adequacy of the translation from the two source language versions available. Such is the adaptability of the technique, that it may also aid the research when both the researcher and available bilinguals are not competent in both languages. An example of such a situation is the study by Fink (cited in Brislin (1970) where an attitude survey needed to be translated from English to the Laotian language. While bilinguals competent in both languages were unavailable, the researcher was able to secure bilinguals who spoke Thai and either English or Lao. A Lao version was eventually derived by going through the following translation procedure: English to Thai to Lao to Thai to English.

c) back-translation is less time-consuming compared to other procedures such as the bilingual technique described below.

Despite the relative ease and consequent widespread use of back-translation, the technique on its own is not foolproof. There are several factors besides good translation that could account for an appearance of equivalence between the source, target and back-translated versions of an instrument. Brislin (1970) lists three such factors:

a) Some bilingual translators may be working with a shared set of rules for translating non-equivalent words and phrases.

b) Some translators may be able to make sense of a poor quality target language version and translate this back to appear like the source version.

c) The first bilingual who translates from the source to target may retain many of the grammatical forms of the source. This may pose no problem to the task of the second bilingual, but would create a target version that would appear exotic to monolinguals of the target version language.

Back-translation is recommended as an initial step to the translation process; however given the possible sources for false-positive results as discussed above, it is not recommended that the technique be used on its own. There are some adaptations of the technique available that may obviate contamination and increase one's confidence that satisfactory results are indeed due to good translation.

6.6.2 Committee approach

The committee approach involves a group of bilinguals who translate from the source to the target language. The benefit of this approach over simple translation which usually involves one person, is that the mistakes of one member can be picked up by others on the committee.

6.6.3 Decentering with multiple iterations

Decentering is the translation technique first described by Werner and Campbell (1970), in which the source language version is open to revision and there is a "de-emphasis of the researcher's language in such a way that the system of symbols supersedes a single culture" (p.399). Decentered translation aims for loyalty of meaning in each language and provides for equal familiarity and colloquialness in the respective languages.

Simple back-translation can be improved by subjecting the instrument to multiple iterations (repeated rounds) of forward and back-translation.

Together, decentering and multiple iterations of back-translation may provide for smoother translation from source to target and allow the researcher greater assurance that the final target version derived is a reliable one, that is equivalent to the original source. The procedure may work in the following manner:

1. Bilingual A translates from Source 1 (original) to Target 1
2. Bilingual B translates Target 1 back to Source 2
3. The researcher compares Source 1 with Source 2 and notes discrepancies. Working with bilinguals A and B the researcher revises the wording of the Source 1 in such a manner that it would allow for a smoother translation. The reworded items constitute Source 3.
4. Bilingual C translates Source 3 to Target 3
5. Bilingual D translates Target 3 back to Source 4
6. Step 3 is repeated and the remaining discrepant items are referred to a new pair of bilinguals for back-translation. The process is continued until all items of an instrument are considered to be equivalent during step 3.

If decentering with multiple iterations is to be used, the researcher has the added advantage in that the source language version of the instrument is also subjected to revision and in a sense, 'centered' closer to the target culture. This may not seem to be of any great value given that the objective of translating the instrument was to establish an equivalent target language version so that original may be dispensed with. However, there are at least two reasons for having a revised, decentered version of the instrument in its original language.

The first emerges as a short-term gain, pertinent only to the task of establishing equivalency between the source and target versions. The bilingual technique (discussed below) of testing translation fidelity involves administering both source and target language versions of an instrument to a group of bilinguals. It is clearly more appropriate to administer a decentered source version which has been revised in such a way that it is more accessible to people of the target group, rather the original source version which may contain terminology and idiomatic expressions that reflect the source culture. In this way, the researcher eliminates the possibility that discrepancies between the responses given by the bilinguals to the two language versions, were due to difficulties of comprehension.

The second benefit of a decentered source version, lies in the long-term application of the instrument. Occasions may arise, albeit seldom, when in a testing situation certain individuals of the target language group indicate a preference for the instrument in the source language. For this case, a decentered source version would be more appropriate than the original. It would more legitimately allow the researcher to group the data from such respondents with that of those who answered the target language version, given that equivalency was shown between the target version and the decentered source version, not the original source version.

6.6.4 Monolingual judges

The use of monolingual judges has been recommended to overcome some of the problems encountered with bilingual translators. Monolingual judges can play a useful role in checking translation quality by ensuring that the target language version is understandable.

6.6.5 Bilingual technique

Once an instrument has been translated both versions are
administered to a sample of bilinguals. Administration is usually
separated by at most, a few days, to avoid any changes in the
respondents life that may affect their responses. The technique
allows for items yielding discrepant responses to be easily
identified.

6.6.6 Pretest procedure

After an instrument has been translated it is recommended that it
is administered to a sample of respondents who typify the
population that the instrument was intended for. This pilot
sample completes the instrument and is asked probe questions
about each item to ensure that future subjects will comprehend
all questions.

Chapter 7

Methodology

The translation of the GHQ into Zulu is a crucial, if not the most important step in this study. All subsequent steps of this study rely heavily on a successful translation. So too would any subsequent study as outlined in the research programme on page 14. The criterion of success is a Zulu version that is semantically equivalent to the original English version, thereby permitting comparisons with results from a wide range of studies. At the same time, this Zulu version would need to be decentered, containing terms and concepts that are familiar in Zulu culture. Translation is a complex process because such objectives are often conflicting and the researcher is forced to make a judicious compromise.

However, as noted in the previous chapter, significant methodological progress has been made in the area of translation and studies such as the present one, can proceed with a useful range of methods and guidelines at its disposal. Notable contributions in this regard have been made in the early works of Werner and Campbell (1970), and Brislin (1970). Although dated, these studies are considered authoritative, with most modern studies making reference to them.

Werner and Campbell (1970) have offered guidelines for writing material in the source language such that it facilitates good translation into a target language. Brislin (1970) has empirically tested the soundness of the Werner and Campbell (1970) recommendations. Although the present study attempts to translate an existing standardised instrument where the source language (English) work is already done, the recommendations of Werner and Campbell (1970) are nevertheless useful when decentering is required and one is called upon to rework the source language in order to facilitate a smooth translation into the target language.

Brislin's (1970) empirical work on back-translation has led to a range of available procedures as well as a seven-step plan for producing good translation. The present translation would be conducted using the back-translation method and would involve a series of steps as recommended by Brislin (1970), referred to as the multi-stage iterative procedure. Decentering would also be employed with the multi-stage back-translation procedure. In addition to such recommendations, the methodology employed has been guided by the work of Chan (1985) and Sriram et al. (1989) in their respective efforts at establishing psychometric equivalence of the Chinese and Kannada versions of the GHQ. The entire method outlined below can therefore be said to involve a combination of steps that have been employed, often separately, by other workers. With some of the revisions described below, particularly Stage 2 (see p75) and stage 6 (see p77), the present method is seen altogether as novel and innovative.

7.1 Preparations for translation

7.1.1 Choice of translators

The quality of translation has been found to be enhanced when the translators are familiar with the content of the material being translated (Brislin, 1970). Beyond being fully bilingual, the translators (especially the English-to-Zulu translator) needed to be familiar with the local dialect or variation of Zulu as it is used in the target community. The teams of translators were therefore chosen with these two criteria in mind. For the first stage of back-translation, both translators held post-graduate qualifications in psychology (one at masters level and the other at honours level, a registered psychometrist). For the second stage, the translators were two teachers involved in teaching Zulu literacy (one a masters student in English and the other an undergraduate). Translators on the panel in stage 4 included all but one of the translators employed in the two back-translation stages, together with two additional persons, one who worked on

translating adult basic education materials into Zulu and the other an experienced adult education trainer. All translators, were fully bilingual in English and Zulu and well acquainted with the Zulu language as it is used in the target community. Only one of the translators was non-African, and would be the only person who could be said to have not experienced South African township life.

7.1.2 Briefing of translation teams

The translators were fully briefed as to the nature of the study and the objectives of their roles. To ensure that translators were working with a common set of objectives and rules, and that they remembered them, they were each provided with a document containing guidelines for their task (See appendix A). This document was constructed largely from recommendation made by Retief (1988) and Werner and Campbell (1970).

Brislin (1986) recommends the use of such guidelines, in order to ensure that translators:
1. have a clear understanding of the original language item;
2. have a high probability of finding a readily available target language equivalent so that they do not have to use convoluted or unfamiliar terms;
3. be able to produce target language items readily understandable by the eventual set of respondents who are part of the data-gathering stage of the research project (Brislin, 1986, p143).

7.2 Back-translations

The plan for translation followed the basic approach outlined on page 69 under the heading "Decentering with multiple iterations". The plan was revised with the inclusion of stage 6 (see below) to stop after two iterations of forward- and back-translation and to allow a panel of translators to work with two sets of Zulu

translations of the GHQ. The benefit of this approach is that
translators can correct each others' errors as recommended by
Minsel et al. (1991). The approach also recognises the point
raised by Triandis et al. (1972), that there is more than one
possible translation that is suitable. This revision was also
influenced by time constraints and the impending departure of two
of the translators. Due to the number of people involved, the
translation steps alone were conducted over a period of two
months.

7.2.1 *Stage 1*

The first translator was given the original English version (see
Appendix B) of the GHQ and asked to translate it into Zulu. The
second translator blindly translated the Zulu-version back into
English. This stage resulted in two English versions of the GHQ
which could be compared. These two versions will be referred to
as the original English and back-translated English,
respectively.

7.2.1 *Stage 2*

A panel of eight evaluators, including the researcher, then
independently compared the back-translated English version with
the original English version and noted items containing
discrepancies in meaning. A form with each item pair was given to
each evaluator. Evaluators rated each item pair on a scale of 1
to 5, where:

1 = totally different meaning
2 = slight variation in meaning - distortion possible
3 = hard to say
4 = quite similar in meaning - distortion unlikely
5 = equivalent in meaning

Evaluators were mostly persons with some postgraduate training, many of them had an English major as part of their qualifications. They were instructed to concentrate on the meanings of items as conveyed in both English versions.

7.2.3 Stage 3

Items that showed semantic non-equivalence in stage 2, that is, those receiving a mean rating of 1 to 3, were referred to a team consisting of the two translators responsible for the translations (Stage 1) and a third bilingual who was experienced with translation.

The team was asked to note the possible sources of error with each of the problem items. There are two possible sources of error. Firstly, the original was not adequately translated into Zulu. Secondly, the original was adequately translated in Zulu, but the back-translation into English was poor. The first source of error is obviously more serious.

Where the problem was due to poor translation into Zulu (source 1), the team was asked to suggest a rewording of the original English, that would facilitate a smoother translation into Zulu ie., decentering.

7.2.4 Stage 4

The GHQ, now with items that were rated as nonequivalent having been decentered, were then given to a new pair of translators for a second iteration of translation into Zulu and then back-translation into English.

7.2.5 *Stage 5*

Discrepancies between the two English versions were once again evaluated by 7 of the people who served on the evaluation panel in stage 2. Items with discrepancies were referred back to the new translation team for identification of the source of error and for rewording of the Zulu version.

7.2.6 *Stage 6*

A panel was formed consisting of the four translators used thus far, and a fifth independent bilingual and the researcher. This panel worked with a list of the items where each item appeared three times, namely;

1. The original English, with some items decentered. This was for reference purposes.
2. The Zulu emerging from the first iteration
3. The Zulu emerging from the second iteration

This panel worked with each item individually. The bilinguals in the team were asked to examine each item and to select the most superior Zulu wording for each item. After much discussion the panel arrived at a selection of the most suitable Zulu translation for each item. For two items the best wording was a combination of the two available Zulu translations. For 22 items the wording from the first translation was deemed to be superior. For 6 items the wording from the second translation was deemed better. The work of this panel produced the final Zulu version that was used in the pretesting step described below.

7.3 Pretest procedure

It has already been noted that bilinguals differ in significant ways from the monolingual target population (see discussion on p46). The bilingual teams of translators were of relatively high education levels in comparison to the intended wider target population. Pretesting as recommended by Pareek and Rao (1980) was considered to be important and was conducted in two stages.

7.3.1 *Stage 7*

The English version, with some items having been rephrased in the decentering steps, was given to a group of three high school teachers of English. They were asked to read the questionnaire and to reword items that they thought would pose a problem in terms of the level of language. They were asked to simplify the English while retaining the original meaning of the item, such that it would be understandable to a standard 8 pupil in their schools.

7.3.2 *Stage 8*

This stage involved pretesting the Zulu version with a monolingual sample. Pretesting is recommended as a useful and important step in translating research instruments (Minsel, Becker and Korchin, 1991; Pareek and Rao, 1980)

The Zulu version of the GHQ that emerged from stage 6 was administered to a sample of six newly literates. The sample consisted of people who were learning English at a literacy class run at the university. People in this group were orally competent in Zulu but would experience difficulty in reading and writing in Zulu. The Zulu version of the GHQ was read to them, item by item. The sample consisted of four females and two males, all between the ages of 30 and 45 years.

They were asked probe questions assessing their comprehension of each item. The following probes were used, "Do you understand this question?" and "Tell me in your own words, what the question is asking?" Difficult items were noted and referred back to the panel of translators.

This stage of the translation testing helped to redress possible biases caused by the use of highly educated translators who are fluent in both English and Zulu. It was conducted to ensure that the final version of the Zulu GHQ that emerged would be applicable to a wide range of Zulu speakers, particularly those who are solely Zulu mother tongue speakers, a characteristic of a large part of the Zulu-speaking population.

Although the GHQ was intended to be used as a self-administered questionnaire, as required in other African sites (Aderibigbe and Gureje, 1992), it may be necessary given high rates of illiteracy in the country to administer the GHQ verbally. This phase was also intended to investigate possible difficulties encountered in verbal administrations of the GHQ. Such fieldtesting was aimed at ensuring that future subjects would comprehend all questions.

Such pretesting confirmed the confidence expressed by the panel in stage 6 with the Zulu version. The pretest sample indicated that all items were understood. Probing for each item showed that the respondents did indeed understand what was being asked by each question. Two items received minor revisions based on the feedback from the pretest sample. This step produced the 'final Zulu' version that was used in the bilingual administration step.

7.4 Bilingual administration

The simplified English version (see Appendix C) and the 'final Zulu' version (see Appendix D) of the GHQ were administered to a group of 257 bilingual students, at three high schools in Pietermaritzburg. The sample (see table 1) consisted of 114 males and 143 females, mean age 18.2 years. The sample consisted of students in the following standards; Std. 8 (56 students), Std. 9 (163 students) and Std. 10 (38 students). Everyone in the sample could be considered to have had almost ten years of schooling and should therefore have been able to understand both the simplified English version and the Zulu version of the questionnaire. Of the total sample of 257, only 9 students indicated that Zulu was not spoken as their home language.

Table 1: Demographic details of the bilingual student sample

	n	%
Sex: Male	114	44.4
Female	143	55.6
Total	257	100
Education:		
Standard 8	56	21.8
Standard 9	163	63.4
Standard 10	38	14.8
Home Language:		
Zulu	246	95.7
English	5	1.9
Other	4	1.6
Not given	2	0.8
Mean Age = 18.2 years		

At each school permission for the study was granted by the
principal. After explaining the purpose of the study and gaining
informed consent, students completed the English version (GHQ-E)
and Zulu version (GHQ-Z) with an average three day interval
between both administrations. On the first occasion, 131
students completed the GHQ-E and 126 students completed the GHQ-
Z. On the second occasion the order of administration was
reversed. The sample used was larger than those employed
elsewhere for such purposes. The Sriram et al. (1989) study used
100 bilingual female college students.

A simplified response format of Yes/No (Yebo/Cha, in Zulu) was
used for each item. This format has been used with other Zulu
versions of the GHQ reported earlier (Leeb, 1986; O'Neil, 1988;
Pillay et al., 1992). The present teams of translators were in
agreement that the original response format of the GHQ was
particularly difficult to convey in Zulu and would ultimately be
confusing for the respondents in both English and Zulu.

The data was coded by the present author and analysed using a
number of programs of the Statistical Package for the Social
Sciences (SPSS).

7.5 Observations from methodology

Butcher and Clark (cited in Brislin, 1986) recommend that
detailed information concerning translation efforts accompany
each item on the final versions of the data-collection
instruments. They reason that as "more and more such information
is made available about translation procedures with existing
tests, future efforts to develop new translations should be far
easier and more psychometrically sound" (Brislin, 1986, p.153).

In addition to the goals advanced by Butcher and Clark above, the present discussion of the results of the back-translation and decentering steps is offered here for three other purposes, namely;

1. There is more than one Zulu version of the GHQ available. Information regarding the translation of individual items will assist potential users of the GHQ in selecting one. Such information will also aid the interpretation of data produced by use of the instrument.

2. Knowledge of the processes involved in arriving at a set of items will enhance any further revision and translation of the GHQ. Translation of other instruments into Zulu, may also benefit from the creation of a pool of terms and psychological concepts that are found to be translatable.

3. A multi-stage design has been employed with the translations. Such reporting will contribute to a body of literature on translation methodology that may elucidate matters on the efficacy of particular translation methods and the relative merits of different steps. Where time is limited, it may help researchers select methods of greatest utility.

Evaluations of the two English versions from the first round of back-translation showed 9 items to have semantic differences. These were items: 7, 31, 35, 39, 47, 54, 55, 56, 58. These items were referred back to the translation team in stage 3. The team were asked to note the source of the discrepancies with each item. Where the problem lay with the translation into Zulu, the team suggested improved rewording. This was the case with three items, namely, items 47, 54 and 55.

Where the problem lay with the original English, in that it was difficult to find an equivalent Zulu phrase, the team reworded the original English (decentering), such that it would allow for a smoother translation. This was the case with the remaining six

items. The English version that was submitted to the next round of back-translation contained these six reworded items.

Items that were decentered for the second stage of back-translation were: 7, 27, 35, 39, 56, 58.

Evaluations of the two English versions from the second round of back-translation showed 6 items to be semantically different. These were items: 14, 26, 32, 33, 35, 39.

Of these 6 items two were repeats in that they had also appeared in the first round of evaluation as possessing semantic non-equivalence. These were items 35, 39. For these items the decentering after the first round of translation had not helped to improve the translation in the second round. Once again, a translation team worked on rephrasing all such problematic items.

The English version used in the bilingual administration step received two types of revision. Firstly, some items were reworded in the decentering process, to allow for smoother translations. Secondly, all items were examined by two high school teachers of English. These teachers suggested rewordings for many of the items in order to bring the English to a level that would be familiar to a standard eight pupil.

It is noteworthy that on the whole the first team of translators produced more superior translations. More items from their translations were selected in stage 6 for the Zulu version presented for pretesting. This is most likely due to the presence of specialist knowledge of psychology and psychological testing within this team. Recall that both translators in the first team were postgraduate students of psychology. Minsel, Becker and Korchin (1991), have noted that translation "errors become less probable if the translation is made by a committee of people who are familiar with the research issue, as well as the different languages" (p.162).

It is also noteworthy that despite the extensive and intensive process of two rounds of back-translation, with evaluation, decentering and rephrasing, two items on the Zulu version still required refinement on the basis of the pretesting with Zulu-speaking monolinguals. This discovery has two crucial implications which have been raised earlier. Firstly translation efforts cannot rely totally on the efforts of bilingual translators. Secondly, pretesting with monolinguals who represent the wider target population is a necessary and important step in testing translation adequacy.

A very interesting observation was noted with item 56 during the translation process. Item 56 failed the test in evaluation 1, passed the test in evaluation 2, yet when the two Zulu versions were compared in stage 6, the translation from the first translation was chosen as superior. This reflects the inherent problems of translation and the suitability of the method employed here. The Zulu in the first translation was good but poorly back-translated into English and therefore failed the test of the evaluators(1). The Zulu in the second translation was inferior to that of first but was back-translated into English in a form that caused it to be passed by the evaluators(2). If one relied solely on the advice of the evaluators, the more superior translation would have been rejected. This example shows the advantage of conducting a full second back-translation, as well as the advantage of a panel of translators checking each other as employed in stage 6.

Chapter 8

Results

Note:

1. The output from SPSS tends to be voluminous. Where appropriate, such output has been provided in various tables. The full output and raw data is available from the Secretary, Psychology Department, University of Natal, Pietermaritzburg.

2. To assist the reader in matching sections of the discussion (Chapter 9) with the relevant sections of the results, the same paragraph numbers have been used for sub-sections of both the results and discussion chapters.

In studies assessing the equivalency of translated versions of the GHQ, a variety of types of analyses have been conducted on data gathered from the bilingual respondents. Studies have examined item-item correlations, item-whole correlations, test-retest correlations, mean scores, endorsement frequencies, reliability coefficients, concordance rates of high scorers and low scorers, and factor structures (Sriram et al., 1989; Chan, 1985; Chan and Chan, 1983).

Several types of analyses were conducted on the GHQ data of the present 257 bilingual respondents. These results are grouped into four broad categories, namely, Total Scores Analyses, Item Analyses, Reliability Analysis and Factor Analysis. For all types of analyses, results are based on the responses to the 30-item GHQ which was administered in both Zulu and English.

Where possible, results are also presented for the 20-item GHQ and 12-item GHQ. The reason for this is two-fold. Firstly, local studies may require a shorter version of the GHQ, particularly if it is to be used as part of a larger battery of instruments or as part of a long questionnaire. Time and personnel constraints within a primary care context may also necessitate the use of a

short screening device. Secondly, no data currently exists for these two shorter versions within a South African context. Other workers may therefore benefit from knowing the reliability indices of these shorter versions with a South African sample. It should however be noted that the 20-item and 12-item GHQ were administered while being embedded within the 30-item version. The results on these two versions presented here cannot therefore be interpreted to be equivalent to situations where these versions are administered on their own, in other words, while not embedded within the 30-item version.

The three versions reported on will henceforth be referred to as the GHQ-30, GHQ-20 and GHQ-12. Furthermore, the items of the GHQ-30 employed here, will be referred to by the numbers given to items by Goldberg (1972) as they appeared in the GHQ-60. It has become a convention in the literature on the GHQ to refer to items by their original GHQ-60 number, even when the shorter versions are employed.

8.1 Total Score Analyses

8.1.1 Mean scores

GHQ-30
Total GHQ scores were calculated for the English and Zulu
versions of the questionnaire completed by each respondent (see
Table 2 and Figure 8.1). The mean score for the English version
was 9.28 (SD = 5.20) and the mean score for the Zulu version was
10.07 (SD = 6.59). On the whole, respondents scored higher on the
Zulu version of the questionnaire. The difference between the
means of the two versions was .79. The overall correlation
between total GHQ scores on the two versions was .73, significant
at P < 0.001. The median score was 8.0 for the English version
and 9.0 for the Zulu version.

GHQ-20
The mean score for the English version was 6.55 and the mean
score for the Zulu version was 7.36 (see Table 2 and Figure
8.1). On the whole respondents scored higher on the Zulu version
of the questionnaire. The difference between the means of the two
versions was less than one (.81). The overall correlation between
total GHQ scores on the two versions was .72 (P < 0.001). For
this shorter GHQ, the median was 6.0 for the English version and
7.0 for the Zulu version.

GHQ-12
The mean score for the English version was 3.54 and the mean
score for the Zulu version was 4.22 (see Table 2 and Figure 8.1).
On the whole respondents scored higher on the Zulu version of the
questionnaire. The difference between the means of the two
versions was less than one (.68). The overall correlation between
total GHQ scores on the two versions was .68 (P < 0.001). On the
GHQ-12, the median was 3.0 for the English version and 4.0 for
the Zulu version.

Table 2: Total scores for the English and Zulu versions of the GHQ-30, GHQ-20 and GHQ-12 (means and correlations)

	English	Zulu	Correlation	Signif.
GHQ-30	9.28	10.07	.73	0.001
GHQ-20	6.55	7.36	.72	0.001
GHQ-12	3.54	4.22	.68	0.001

Figure 8.1 :
Mean scores on the English and Zulu versions of the GHQ-30, GHQ-20 and GHQ-12.

Mean score
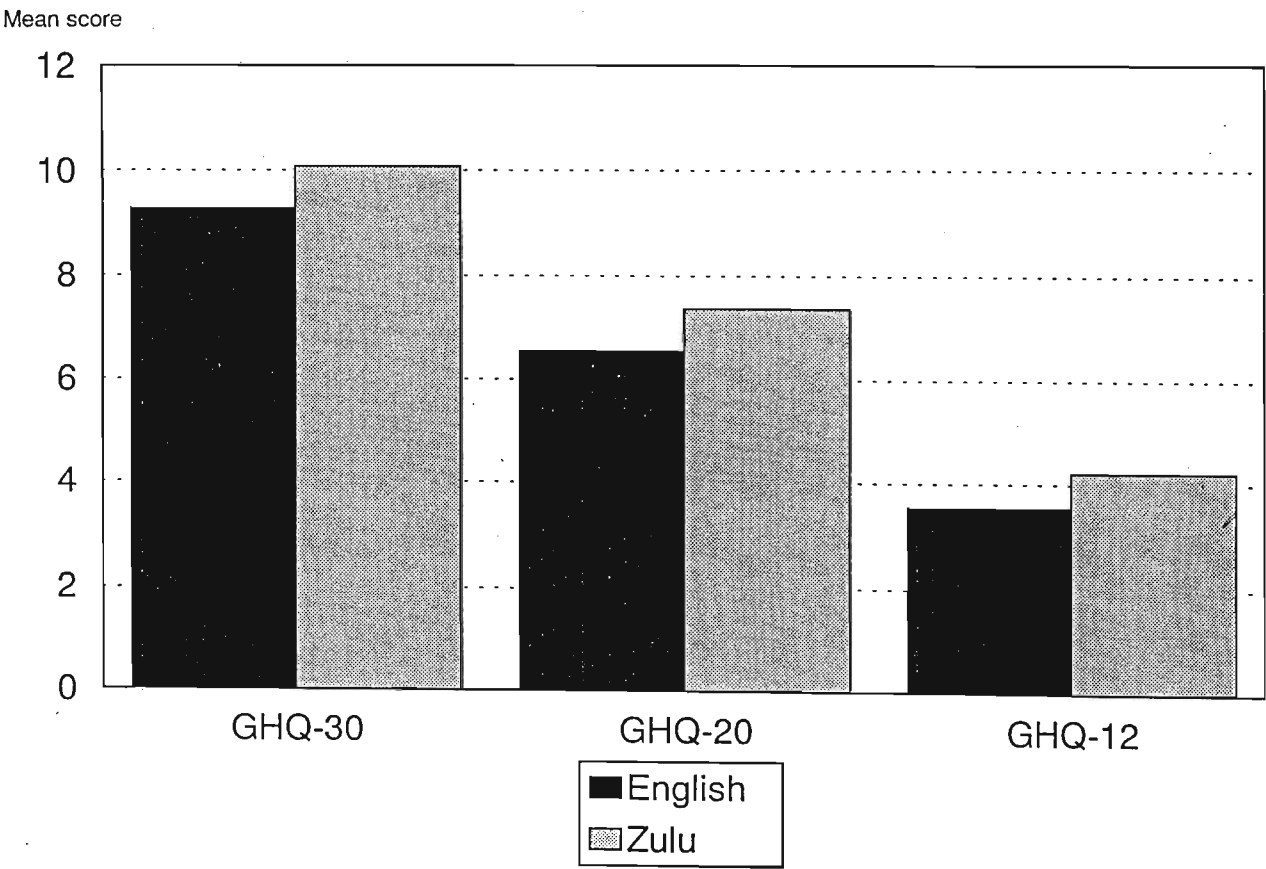
8.1.2 Test-retest equivalency

The sets of total scores for each subject were separated into two administration groups, namely English-Zulu and Zulu-English. The correlations of total scores achieved magnitudes of .66 and .78 for English-Zulu and Zulu-English test-retest situations. As stated earlier the overall test-retest correlation was .73. All three correlations were significant at $P < 0.001$.

8.1.3 Concordance between high scorers

With the cutoff set at 8 (median for the English version), there are 127 high scorers on the English version and 103 high scorers on the Zulu version. The concordance rate of high scorers is 81%.

8.1.4 Effects of administration order

To check whether the order of administration had any influence on the results, total GHQ scores on the English and Zulu versions where examined for first and second administrations separately. The subject pool was divided into two groups (see Table 3).

Group A (N = 131) received the English version first. The mean GHQ score for Group A was 8.56 on the English questionnaire and 8.24 on the Zulu questionnaire.

Group B (N = 126) received the Zulu version first, the mean score for Zulu was 11.96 and for English the mean score was 10.02.

Figure 8.2 illustrates the comparative means on the English and Zulu versions, achieved by Group A and Group B.

Table 3: Total GHQ scores for each administration group (means and correlations)

Administration	English version	Zulu version	Correlation	Signif.
Group A (English 1st)	8.56	8.24	.66	.001
Group B (Zulu 1st)	10.02	11.96	.78	.001

Figure 8.2 :
Mean scores on the English and Zulu versions for each administration group.
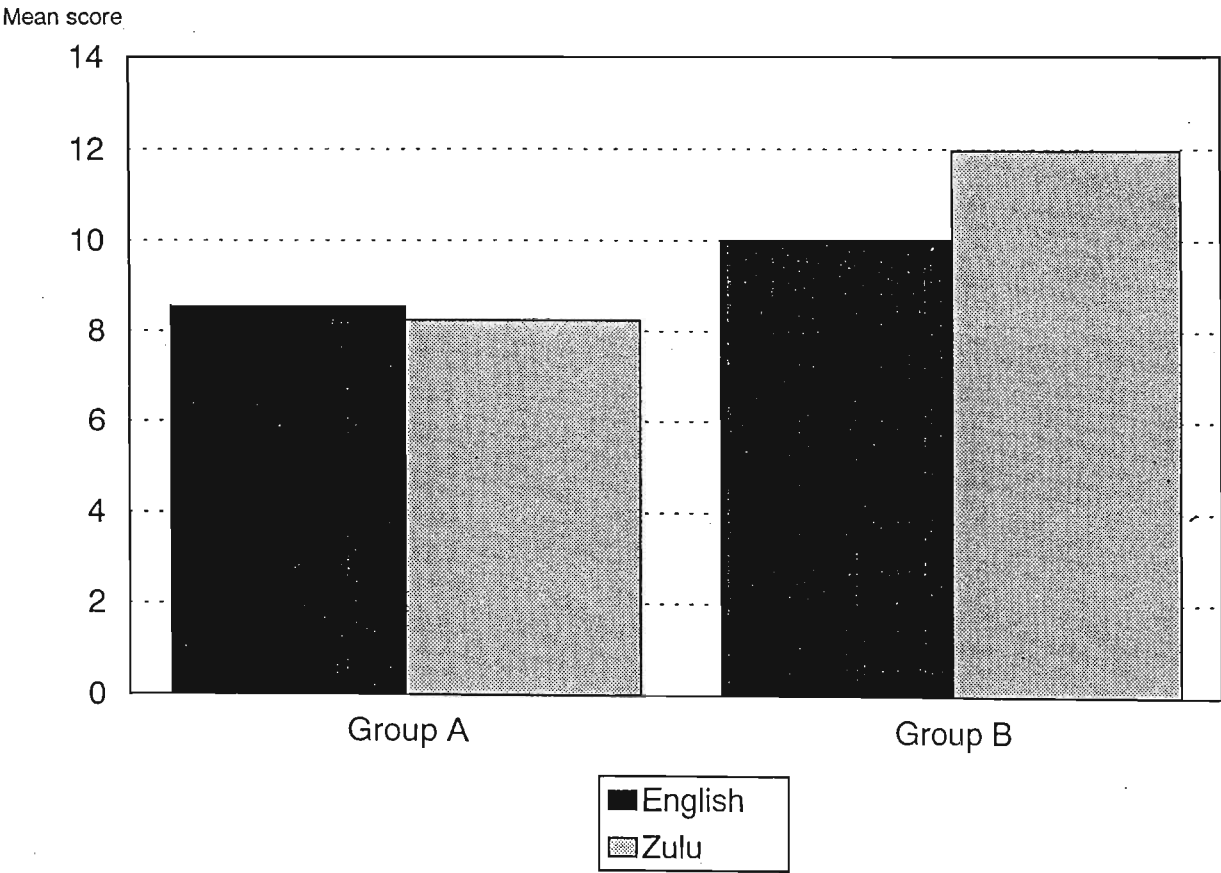
8.2 Item Analyses

8.2.1 Endorsement frequencies

Endorsement frequencies of the items, that is the proportion of respondents who endorsed the item in the pathological direction, were computed separately for the English and Zulu versions (see Table 4). For the English version, the range of endorsement frequencies was .150 - .678. The range for the Zulu version was .130 - .609. Sixteen items received higher endorsement rates in the Zulu version and the remaining 14 items received higher endorsement rates in the English version.

For each item, the endorsement frequency on the English version was compared with the respective endorsement frequency on the Zulu version. The difference between each pair of endorsement frequencies constitutes the endorsement discrepancy for that item. Table 5 shows the nine items (items 7, 30, 35, 40, 43, 49, 41, 55, 58) which possessed the greatest endorsement discrepancies (greater than the magnitude of 0.10, between the two versions. Item 35 had the largest difference (.289) and was the only item to possess a discrepancy of a magnitude greater than .20.

Of the nine items with high endorsement discrepancies, 8 items also appear in the GHQ-20 and four items appear in the GHQ-12.

Table 4: Endorsement frequencies for the English and Zulu versions of the GHQ-30.

GHQ Item	English version	Zulu version	Discrepancy
7	.178	.353	.175
14	.331	.404	.073
21	.199	.223	.024
26	.544	.492	.062
27	.328	.236	.092
30	.259	.412	.153
31	.174	.224	.050
32	.161	.133	.028
20	.270	.337	.067
33	.294	.208	.086
35	.419	.130	.289
36	.184	.248	.064
28	.303	.254	.049
39	.252	.344	.092
40	.360	.559	.199
42	.220	.297	.077
43	.678	.529	.149
53	.180	.157	.023
45	.371	.368	.003
46	.150	.219	.069
47	.426	.421	.005
49	.491	.609	.118
41	.326	.523	.197
50	.392	.383	.009
51	.249	.349	.100
52	.344	.302	.042
54	.345	.338	.007
55	.256	.364	.108
58	.417	.527	.110
56	.309	.247	.062

Table 5:

Items with endorsement discrepancy greater than .10

(endorsement frequencies for English and Zulu are displayed)

	English	Zulu
7 Have you recently been able to pay attention to whatever you're doing?	.178	.353
30 Have you recently been happy with the way you have done your work?	.259	.412
35 Have you recently felt that you are playing a useful part in things?	.419	.130
40 Have you recently felt you could not solve your problems?	.360	.559
41 Have you recently been finding life is a problem all the time?	.326	.523
43 Have you recently been finding things hard?	.678	.529
49 Have you recently been feeling unhappy and depressed?	.491	.609
55 Have you recently been feeling nervous and worried all the time?	.256	.364
58 Have you recently found at times you couldn't do anything because you felt worried?	.417	.527

8.2.2 Item-total correlations

Item-total correlations were computed separately for the English
and Zulu versions (see Table 6). For the English version, item-
total correlations ranged from -.0190 to .5528. For the Zulu
version, the range was .0818 to .6056. One item in the English
version, namely, item 35 had a negative item-total correlation.
All item-total correlations for the Zulu version were positive.

8.2.3 Item-item correlations

The response to each English item was correlated with those of
the corresponding Zulu item across all respondents. The resulting
item-item correlations (see Table 7) were all positive and ranged
from .13 to .52. Significant correlations were obtained for 28 of
the 30 items. Item 43 and item 54 were the only two items that
did not obtain significant item-item correlations. Both items 43
and 54 appear in the GHQ-20 and item 54 also appears in the GHQ-
12. Of the items for which correlation attained significance, 26
were significant at $P < 0.001$ and the remaining two items were
significant at $P < 0.01$.

Table 6: Item-total correlations for the English and Zulu versions of the GHQ-30.

GHQ Item	English version Item-total Correlations	Zulu version Item-total Correlations
7	.2744	.2846
14	.4024	.4381
21	.0758	.2502
26	.0962	.0818
27	.1778	.5464
30	.3359	.4237
31	.2149	.4429
32	.2727	.4591
20	.3770	.3872
33	.2943	.0995
35	-.0190	.3076
36	.2372	.3420
28	.4347	.3525
39	.5527	.5570
40	.4424	.3990
42	.4733	.5586
43	.3697	.3921
53	.1829	.5026
45	.3453	.4379
46	.3537	.4882
47	.3309	.6046
49	.3793	.4888
41	.4992	.4420
50	.2784	.4429
51	.3826	.5719
52	.2632	.6056
54	.1996	.4487
55	.5528	.6056
58	.4007	.5748
56	.3303	.5802

Table 7: Item-item correlations for the GHQ-30

GHQ Item	Item-item Correlation	Significance
7	.3525	.001
14	.5210	.001
21	.4383	.001
	.3262	.001
27	.2846	.001
30	.4388	.001
31	.4038	.001
32	.2447	.001
20	.2203	.001
33	.2705	.001
35	.1703	.01
36	.1958	.001
28	.3307	.001
39	.4593	.001
40	.2998	.001
42	.4007	.001
43	.1247	NS
53	.2469	.001
45	.2537	.001
46	.1512	.01
47	.2082	.001
49	.3482	.001
41	.2666	.001
50	.2494	.001
51	.3721	.001
52	.1957	.001
54	.1417	NS
55	.2418	.001
58	.2813	.001
56	.3251	.001

8.3 Reliability Analyses

The internal consistency of the GHQ was computed separately for the two language versions. Reliability coefficients are presented in Table 8. The resulting alpha coefficients were .81 for the English version and .89 for the translated Zulu version.

For the GHQ-20, coefficients of .75 and .84 were obtained for the English and Zulu versions respectively. For the GHQ-12, resulting coefficients were .64 and .77 for English and Zulu versions respectively.

Table 8:
Reliability coefficients for English and Zulu versions of the GHQ-30, GHQ-20 and GHQ-12.

GHQ version	Alpha coefficient
English GHQ-30	.81
Zulu GHQ-30	.89
English GHQ-20	.75
Zulu GHQ-20	.84
English GHQ-12	.64
Zulu GHQ-12	.77

8.4 Factor Analysis

Factor analysis was conducted for the English and Zulu versions
separately. As shown in Table 9, for the English version ten
factors emerged with eigenvalues greater than 1, accounting for
57.5% of the total variance. For the Zulu version, eight factors
emerged with eigenvalues greater than 1, accounting for 57.4% of
the variance.

Table 9:
Factors on the English and Zulu versions with eigenvalues greater
than 1.

| Factor | English version | | | Zulu version | | |
	Eigen-value	% of variance	Cum. var.	Eigen-value	% of variance	Cum. var.
1	5.20	17.3	17.3	7.76	25.9	25.9
2	1.92	6.4	23.7	1.9	6.3	32.2
3	1.59	5.3	29.0	1.5	4.9	37.1
4	1.48	4.9	33.9	1.5	4.8	41.9
5	1.34	4.5	38.4	1.3	4.3	46.2
6	1.29	4.3	42.7	1.2	3.9	50.1
7	1.19	4.0	46.7	1.1	3.8	53.9
8	1.16	3.9	50.5	1.1	3.5	57.4
9	1.09	3.6	54.2			
10	1.01	3.4	57.5			

For both the English and Zulu versions the first factors had
eigenvalues (5.20 and 7.76 respectively) which were much larger
than subsequent factors. The first factor for the English version
accounted for 17.3 % of the variance. The first factor for the
Zulu version accounted for 25.9% of the variance. All subsequent
factors on both versions accounted for less than 6.5% of the
variance. Figure 8.3 depicts the decline in variance accounted
for by factors subsequent to the first.

Figure 8.3: Percentage of variance accounted for by the first 10 factors on the English version and first 8 factors on the Zulu version

% variance
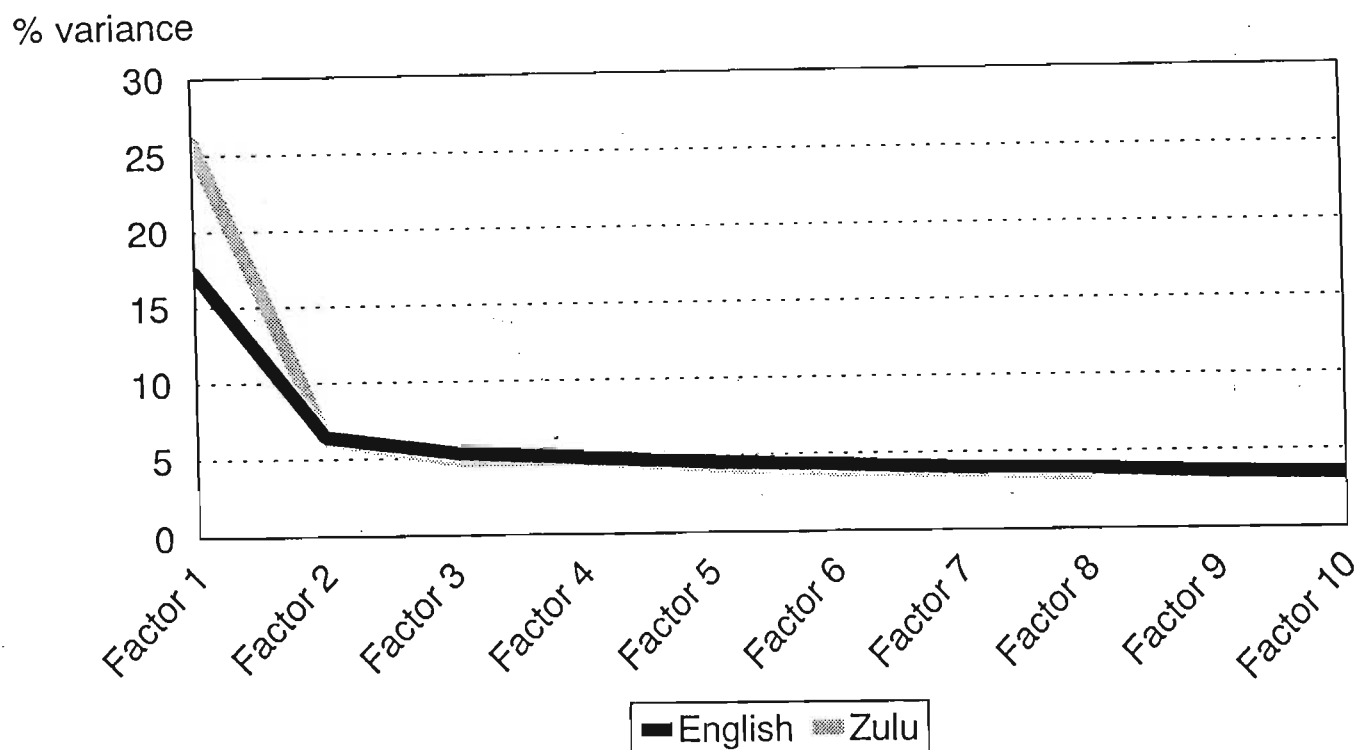
The first factors of each version were examined in the unrotated matrixes with respect to items that loaded heavily on these first factors (Table 10). It was decided to consider only items with loadings greater than .40 for this analysis. Other examinations of factor solutions of the GHQ have used loading cut-offs ranging from a magnitude of .30 (Chan and Chan, 1983) to .40 (Medina-Mora et al., 1983). Factor 1 of the English version contained sixteen items with high loadings. Factor 1 of the Zulu version contained twenty three items with high loadings. There were 15 items that had high loadings on both the English and Zulu versions, portraying the GHQ as more unidimensional.

Table 10:
Factor 1 on the English and Zulu versions (unrotated matrix)
Items with loadings greater than .40

GHQ item	English Factor 1	Zulu Factor 1
14	.51	.50
27		-.60
30	-.41	-.47
31		-.48
32		-.51
20	.47	.43
28	-.51	
39	.67	.61
40	.55	.44
42	-.55	-.60
43	.44	.44
53		-.55
45	.43	.50
46	-.41	-.53
47	.44	.67
49	.48	.54
41	.61	.51
50		.51
51	.46	.65
52		.69
54		-.50
55	.65	.69
58	.51	.63
56		.65

Given the substantial decline in percentages of variance which factors subsequent to the first accounted for (illustrated in Figure 8.3), it was decided to enter the first six factors of each version into a rotated matrix for interpretation. The scree plot indicated that the first six factor were enough to retain for rotation. Varimax, an orthogonal rotation was used with these factors. Layton and Rust (1986) note that the number of factors chosen for rotation in various studies of the GHQ range from 4 to 6. They further note that the number of factors selected for rotation is rather arbitrary and dependent on the relative size of the first factor. In their factor analysis of the GHQ, Layton and Rust (1986) chose 5 factors for rotation from a list of 17 factors with eigenvalues greater than 1. As with the present study, Chan (1985) examined 6-factor solutions of the English and Chinese versions of the GHQ.

The six factors retained for rotation on the English version individually accounted for 17.3%, 6.4%, 5.3%, 4.9%, 4.5% and 4.3% of the variance. Together they accounted for 42.7% of the total variance on the English version. The six factors retained for rotation on the Zulu version individually accounted for 25.9%, 6.3%, 4.9%, 4.8%, 4.3% and 3.9% of the variance, cumulatively they accounted for 50.1% of the total variance.

Once again items with loadings greater than .40 on each factor on the rotated matrix were selected for further examination and are presented in Table 11. As it turned out, most item loadings were in excess of .45. From the rotated matrixes, eight items loaded heavily on the first factor of the English version, while six items emerged with heavy loadings on the first factor of the Zulu version.

Table 11: The six-factor solutions of the English and Zulu versions (Items with loadings greater than .40).

English version

Factor 1
14 lost much sleep because of worry?   .65
20 been having restless, disturbed nights?   .56
39 been feeling stressed and worried all the time   .64
40 felt you could not solve your problems?   .50
49 been feeling unhappy and depressed?   .51
41 been finding life is a problem all the time?   .54
55 been feeling nervous and worried all the time?   .62
58 found at times you couldn't do anything because you felt worried?   .55

Factor 2
45 been getting scared or nervous for no good reason?   .50
50 been feeling unsure of yourself?   .44
51 been thinking of yourself as a useless person?   .59
52 felt that life is mostly hopeless?   .63
56 felt that life is not worth living?   .53

Factor 3
7 been able to pay attention to whatever you're doing?   .64
30 been happy with the way you have done your work?   .55
31 been able to feel love and affection for those close to you?   .47
32 been able to get on with other people?   .42
42 been able to enjoy your normal daily activities?   .45
46 been able to face your problems?   .46

Factor 4
21 been able to keep yourself busy and occupied?   .44
36 felt able to make decisions about things?   .53
28 felt you were doing most things well?   .45
43 been finding things hard?   -.45
54 been feeling reasonably happy, overall?   .59

Factor 5
26 been getting out of the house as much as usual?   .51
33 spent much time talking with people?   .70
50 been feeling unsure of yourself?   .46

Factor 6
35 felt that you are playing a useful part in things?   .58
53 been feeling hopeful about your own future?   .60

Table 11 (continued)

Zulu version

Factor 1

50 been feeling unsure of yourself?	.59
51 been thinking of yourself as a useless person?	.69
52 felt that life is mostly hopeless?	.88
55 been feeling nervous and worried all the time?	.88
58 found at times you couldn't do anything because you felt worried?	.45
56 felt that life is not worth living?	.55

Factor 2

30 been happy with the way you have done your work?	-.42
28 felt you were doing most things well?	-.58
39 been feeling stressed and worried all the time	.52
40 felt you could not solve your problems?	.49
43 been finding things hard?	.46
47 found everything going against you?	.43
49 been feeling unhappy and depressed?	.58
54 been feeling reasonably happy, overall?	-.43
58 found at times you couldn't do anything because you felt worried?	.47

Factor 3

7 been able to pay attention to whatever you're doing?	.58
21 been able to keep yourself busy and occupied?	.62
27 been managing as well as most people would in your situation?	.42
30 been happy with the way you have done your work?	.50
31 been able to feel love and affection for those close to you?	.43
36 felt able to make decisions about things?	.51
42 been able to enjoy your normal daily activities?	.42
46 been able to face your problems?	.50

Factor 4

32 been able to get on with other people?	-.69
35 felt that you are playing a useful part in things?	-.56
39 been feeling stressed and worried all the time?	.41
53 been feeling hopeful about your own future?	-.45
47 found everything going against you?	.47
41 been finding life is a problem all the time?	.53

Factor 5

14 lost much sleep because of worry?	.68
20 been having restless, disturbed nights?	.61
36 felt able to make decisions about things?	-.53

Factor 6

26 been getting out of the house as much as usual	.44
33 spent much time talking with people?	.61

Chapter 9

Discussion of Results

9.1 Total Score Analyses

9.1.1 Mean Scores

If the Zulu version of the GHQ is equivalent to the English version, respondents should produce very similar scores on the two versions. Variations in a respondent's two scores could be created by a significant change in the respondent's condition or by retest effects. The first possible cause has been minimised via a methodological choice. Respondents answered both versions within a short space of time. Furthermore, all GHQ questions ask the respondent to compare their 'recent past' with the way they usually feel. Both administrations should fall within the period of their 'recent past'. The sample used (n=257) was also large enough to absorb any possible significant changes that may have occurred in the lives of, most likely, a few respondents. The second possible cause of retest effects should also be minimised by the present methodological considerations. Roughly half the respondents received the English version first and the other half received the Zulu version first.

If these two possible causes are rendered unlikely to influence any variation in the respondents' scores then any variations that are observed would more likely be attributable to translation problems and a state of non-equivalence between the two versions. The mean scores for the two versions of the GHQ-30 are very similar (9.28 and 10.07). The difference in the mean scores (.79) for the two versions is less than one point. With a difference in means of such low magnitude, as illustrated in Figure 8.1, the two versions can be said to be acceptably equivalent in measurement terms.

The overall correlation between total GHQ-30 scores on the two versions was .73 (P < 0.001). This was lower than the correlation coefficient of .81 found between the total scores on the English and Kannada versions (Sriram et al., 1989), but closer to the correlation coefficient of .78 between the English and Chinese versions. The correlation between total scores on the English and Zulu versions was significant (P < 0.001) and is of an acceptable level. This result points to the existence of comparable forms between the English and Zulu versions.

Correlation coefficients became smaller with the shorter versions. This trend has been reported with other attempts that have compared the total scores of the English version with a translated version. The overall correlation between total GHQ-20 scores on the two versions was .72 (P < 0.001). For the GHQ-12, the overall correlation between total scores on the two versions was .68 (P < 0.001). Here too, this is lower than the correlation coefficient of .76 found between the total scores on the 12-item English and Kannada versions (Sriram et al., 1989). It is noteworthy that the correlation coefficients of the GHQ-20 and the GHQ-12 also reached significance at p < 0.001.

As noted, the difference between the mean GHQ scores for the English and Zulu versions was less than 1. This applied to the GHQ-30 (.79), as well as the GHQ-20 (.81) and GHQ-12 (.68). A difference in means of such low magnitude is highly desirable when establishing equivalence between two language versions of an instrument.

In the development studies of the GHQ, Goldberg (1972) used a cutting-score of 4 or 5 for the GHQ-30. People scoring above this were considered to be a 'case'. If Goldberg's cutting-score was employed for the present study, the large majority of respondents would be classified as quite severely distressed on both the English and Zulu versions. About 75 % of the sample scored over 5 on the English version and 72 % scored over 5 on the Zulu version. With this cutting-score, high rates of 'caseness' would

also be found with the respondents in the three South African studies that employed the 30-item GHQ, namely, the Leeb (1986), O'Neil (1988) and Pillay et al. (1992) studies.

O'Neil (1988) found a mean GHQ score of 10.00 (SD=5.16) using the 30-item version with a sample of South African shop stewards. This is almost identical to the results produced by the Zulu version in the present study. Although the Leeb (1986) study does not report a mean GHQ score, it is evident that 71.7% of her sample (N=113) of employed and unemployed adults, scored between 1 and 10. Pillay et al. (1992), also do not report mean GHQ scores but state that as many as 94,4 % of their sample of primary caregivers from a conflict-ridden area, scored above 4.

Goldberg (1972) suggests a cutting-score of 3 or 4 for the 20-item version of the GHQ. With the present sample, 69% on the English version and 68% on the Zulu version would be regarded as a case using Goldberg's cutting score of 4. Once again, a large percentage of the sample would be categorised as cases using Goldberg's cutoff.

For the 12-item version, Goldberg suggests a cutting-score of 1 or 2. Applying this threshold (2) to the present study would create a case-rate of 63% of the sample with the English version and 66% with the Zulu version. These case-rates are slightly higher than those reported by Turton et al. (1990) who used the GHQ-12 with standard 10 pupils in Alexandra and reported that 59% of the sample scored more than 2.

The present study does not allow for appropriate cutting-scores to be determined. However, on the basis of the mean GHQ scores of the present study and judging from the results of all the local studies, it would seem that South African samples tend to score higher on the GHQ than their western counterparts. This does not however necessarily mean that the South African samples are more distressed. It is quite likely that the cut-offs used for western samples are too low.

Two South African studies, namely, Pillay et al. (1992) and
Turton et al. (1991) used the various cutting scores recommended
by Goldberg (1972). Leeb (1986) found that her scores were high
and therefore divided her sample into three groups on the basis
of their total GHQ scores on a Zulu version:

Group 1 (Scores 1 - 10) - 'normal'
Group 2 (Scores 11 - 20) - 'moderately distressed'
Group 3 (Scores 21 - 30) - 'severely distressed'

In terms of Leeb's categorisation, the bulk (62%) of the present
respondents of the Zulu version would be considered to be
'normal'. A further 29% of the present sample would be regarded
as 'moderately distressed' and the remaining 9% would be
considered 'severely distressed'.

It should however be noted that Leeb's (1986) categorisation did
not emerge from an attempt at concurrent validity as it is not
based on any independent data. Goldberg's (1972) cutting scores
were all born out of extensive work in which GHQ scores were
compared with the opinion derived from standardised clinical
interviews. The resultant cutting-scores were thus guided by the
goals of maximising sensitivity and specificity and minimising
the overall misclassification rate. As discussed in the section
on 'Sensitivity and specificity' (p32) considerable attention has
been given to setting appropriate cutting scores for use of the
original English version of the GHQ. Applications with translated
versions of the GHQ have also paid attention to this, usually
once translation equivalence has been demonstrated (Chan, 1985).

Likewise, the use of the GHQ in this country for case screening
would require appropriate cutting scores to be determined. Proper
norming studies involving random samples and studies which
validate the Zulu version of the GHQ against other measures,
preferably full standardised clinical interviews, are required in
order to arrive at an appropriate cutting score for South African
samples. As argued in the section entitled "A proposed research

programme for instruments of foreign origin" (p14) norming and validity studies will allow for the development of a proper standardised Zulu-version of the GHQ.

9.1.2 Test-retest

All test-retest correlations in the present study were significant. This indicates a level of stability in the repeated measures of GHQ. The test-retest correlations here were however, lower than those observed between the English and Chinese versions (Chan, 1985) where English-Chinese, Chinese-English and overall correlation was .77, .87 and .78 respectively. The English-Zulu test-retest correlation (.66) was lower than the overall test-retest correlation (.73), while the Zulu-English correlation exceeded that of the overall correlation. This appears to follow the same pattern observed with the English and Chinese versions.

9.1.3 Concordance rate for high-scorers

It is interesting that there were more high scorers on the English version given that the Zulu version had a higher mean than the English version. It would seem that the mean on the Zulu version was inflated by outliers, that is, respondents who scored extremely high on the Zulu version. The standard deviation for the Zulu version was 6.59 as opposed to 5.20 on the English version.

9.1.4 Effects of administration order

For both groups A and B (Table 3), there is a decline in GHQ
scores from first to second administrations. Longitudinal studies
involving use of the GHQ have shown a retest effect, where lower
scores have been received from samples on the second and
subsequent administrations (Ormel, et al., 1989). The present
results show a similar retest effect, a decline in scores
irrespective of which language is administered first.

Although they are in the same direction, the declines are however
not of the same magnitude, and seem to be influenced by a
language factor. The drop from English to Zulu is an average .32,
while the drop from Zulu to English, is a much larger 1.94.

A possible explanation for this may lie in the familiarity and
comfortableness experienced by the present sample with the two
languages involved. The large majority of the respondents (246)
speak Zulu as their first language. It would be safe to say that
Zulu is the language that respondents were most comfortable and
familiar with. Of the two languages, English is their second
language and could thus be considered their 'non-native'
language.

As discussed earlier (p47) bilinguals have been found to offer
different responses to the same question when asked in each of
their two languages. A number of explanations have been posited
for such discrepancies. One such explanation, the social
desirability hypothesis claims that bilinguals engage in a type
of impression management and therefore attempt to present
themselves in a more socially desirable manner when responding to
questions that are presented in their second or non-native
language (Tyson et al., 1988; Marin et al., 1983).

English is a second language for the large majority of the
respondents in the present sample. The present respondents could
therefore be scoring lower on the English version of the GHQ in

order to present themselves as more 'normal' which would be more socially desirable. It has been argued that social desirability effects are likely to occur when there is a status difference between the two cultures in question (Tyson et al., 1988). Years of apartheid has provided ample grounds for status differences to exist between the predominantly English-speaking and Zulu-speaking population groups in this country. Some evidence of this is provided in the section entitled "The South African Context" (p4).

An alternative explanation of language acquisition effects (see p49) could also be an influencing factor in the discrepancies found in the present results. Virtually everyone in the present sample would be categorised as coordinate bilinguals. They would therefore have different meaning systems when answering the questions in English and Zulu. According to this explanation, discrepant responses are to be expected.

A combination of the social desirability hypothesis and the retest effect would provide a reasonable basis for understanding the 'language/order-of-administration' effect evident in the present results. With Group A (English first) respondents score relatively low on their first occasion because it is their second language and they feel the need to present a more socially desirable image. When they get their Zulu GHQ they score lower than they did on the English version because of the retest effect, but only marginally lower (.32) because it is the language that they are most comfortable with.

With group B, who receive the Zulu GHQ first, respondents score high on their first occasion because they are responding in the language that they are most comfortable with and which evokes less of a need for impression management. When they receive the English version, their decline in scores (1.94) is pronounced by both the movement to their second language and the retest effect.

It is also worth noting that because one of the main objectives of this study is to produce an adequately translated Zulu version of the GHQ, considerably more effort was put into ensuring that the Zulu was simple and accessible. By comparison, less work was done in ensuring that the English version was also accessible. The reliability coefficient for the English version is lower than that of the Zulu version. It is possible that these differences between the two language versions could also have influenced the effect discussed above.

However, despite the decline in GHQ scores from first to second administrations and the difference in scores between the two language versions both the English-Zulu and Zulu-English retest correlations, .66 and .78 respectively, were significant at $P < 0.001$.

Repeated administrations of the GHQ to bilinguals appears to be a common method for assessing equivalency. Both Chan (1985) and Sriram et al. (1989) employed this method. However the effects of administration order seem to have been overlooked in both these studies. Given the present results, examinations of such effects seem worthwhile when trying to demonstrate the psychometric properties of an instrument.

9.2 Item Analyses

9.2.1 Endorsement frequencies

Examining the discrepancies in endorsement frequencies between the two versions is one indicator of the level of equivalence between them. High equivalency between the two versions can be demonstrated by discrepancies in endorsements that are of low magnitudes.

For most of the items, the endorsement discrepancies are small. The ranges of endorsement frequencies are similarly large for both the English version (.150 - .678) and the Zulu version (.130 - .609). However the endorsements on items are mostly concentrated around the minimum values on both ranges.

Studies examining the endorsement frequencies between the English and translated versions of the GHQ have used a discrepancy magnitude of 0.10 as a cut-off for separating items of high discrepancy from those of low discrepancy (Sriram et al., 1989; Chan, 1985) Nine items in the present study fall above the cut-off of 0.10 (Table 5). This is too large a number of items for the 30-item GHQ. Chan (1985), comparing the English and Chinese versions found only seven items with high endorsement discrepancy on the 60-item version. One mitigating factor, which reflects on the level of equivalency between the English and Zulu versions, is that only one item (item 35) had a discrepancy of a magnitude greater than .20.

Six of the nine items with discrepancies greater than 0.10 (items 7, 30, 35, 40, 43, 58) tap an aspect of behaviour that one may refer to as 'not coping with tasks'. The magnitude of the discrepancy with these items arose because four of them (items 7, 30, 40, 58) received much higher endorsements in Zulu than in English. For the other two items (items 35, 43) the greater endorsements occurred in the English version.

The remaining three items with discrepancies greater than 0.10 (items 41, 49, 55) deal with feelings of 'dissatisfaction, unhappiness and depression'. With each of these items the discrepancy arose because of much higher endorsement in the Zulu version.

The large number of items affected by discrepancy in endorsement frequencies warrants further discussion and an investigation into possible causes. Discrepancies could occur if a word or phrase was badly translated or was vague in either of the language

versions. An examination of the wording of these items reveals two sets of terms that are repeated or common to more than one item. Three items (items 41, 55, 58) have one word in common, namely "isikathi" in Zulu or "time" in English. It is possible that this word created different meanings for these items in the different language versions. This appears to be a likely explanation if one takes into consideration item 39 which also contains the word "time". The difference with Item 39, which does not belong to this set of items with discrepancies greater than 0.1. is that the word "time" appears in the English version but does not appear as "isikathi" in the Zulu version. Post-analysis discussion with a translator revealed that the phrase "sonke isikathi" in item 41 and item 55 may have cued the respondents into thinking about a "particular period" rather than "all the time" as referred to in the English version of these two items. It is quite likely that this difference inflated the endorsement discrepancy for these to items. Given the discrepancy, the translator suggested that the Zulu term "njalo" (i.e. always) may be a more suitable replacement for the phrase "sonke isikhati". Further research would be necessary to establish the adequacy of rephrasing.

Items 35 and 43 both use the word "things" in English and "izinto" in Zulu. A similar problem as with "isikathi-time" could be at play here. In post-analysis discussions with the translator it was suggested that the term "izinto" may have cued the respondents into thinking about "material things" or "objects". In the English version, the term "things" is used in a more general sense. Here also, possible rephrasing and evaluation may be required. The problem with item 35 is further explored in the later discussion on item-total correlations.

Noteworthy of consideration, is the fact there is a roughly equal split in terms of in which version higher endorsements were made. There were 16 items that received higher endorsement rates in the Zulu version and 14 items which received higher endorsement rates in the English version. It would be problematic if the higher

endorsements were concentrated in one version only. Such a situation would create doubts regarding the quality of translation and could possibly point to a contamination attributable to influence of the language of questioning.

The micro-linguistic analysis attempted above could explain the cause for the high discrepancies with only five of the nine items. In general however, it must be recalled that great lengths were taken to ensure that the translated Zulu was clear and understandable. This was seen as necessary because of the objective of wanting to establish a Zulu version that would be used on its own. By comparison, less effort was put into the English version. It is possible that discrepant endorsement frequencies between the English and Zulu version arose because of the respondents' differing level of competency with the two languages. However, despite this possibility, only one (Item 43) of the nine items with discrepant endorsement frequencies failed to produce a significant item-item correlation (see section on item-item correlations below). The discrepancy in endorsement of Item 43 arose because of substantially higher endorsement of the item in the English version.

9.2.2 Item-total correlations

In the development of unidimensional scales, examining item-total correlations is a means of determining the internal consistency of an instrument. Anastasi (1982) notes that during test construction only those items with significant item-whole correlations are retained. A test whose items are selected in such a manner is said to possess internal consistency. Via this method, a certain level of confidence is gained in the items because each item is shown to differentiate among the respondents in the same direction as the whole instrument does.

In the Leeb (1986) study item-total correlations for a Zulu version GHQ-30 ranged from -.26 to .59. A far more acceptable range of item-total correlations (.08 to .61) has been obtained with the Zulu version employed in the present study. Of even greater importance, is the absence of any item-total correlations or zero or negative value in the Zulu version of the GHQ used in this study. This is a most satisfactory condition, one that has not been achieved with other Zulu versions of the GHQ. Leeb (1986) reported as many as seven items (items 7, 42, 46, 54, 27, 31, 53.) which all correlated negatively with the total and a further two items (items 35, 36) that are close to zero correlation. O'Neil (1988) reported two items (items 36, 21) which correlated negatively with the total.

Leeb (1986) notes that the presence of items with negative or zero item-total correlations could be because of poor phrasing of the items or because the item is irrelevant. It has been noted earlier that the translation procedure employed by Leeb (1986) was somewhat limited. The absence of items with negative or zero item-total correlations in the present study would be an indication of the value and success of the careful and rigorous translation procedure that was employed. It also indicates an internal coherence of the instrument with all items being relevant to the sample employed.

The only item with a negative item-total correlation in the English version of the instrument employed was item 35, "Have you recently felt that you are playing a useful part in things". It is possible that item 35 may be assessing a different aspect compared to the rest of the items in the instrument. However, in the discussion on endorsement frequencies, item 35 has already been identified as the item with the largest discrepancy in the instrument. It was the only item with an endorsement discrepancy larger than 0.2. It was also noted that the discrepancy arose because of much higher endorsement of the item in the English version. A explanation for the large discrepancy seemed to lie in the possibly different way the words "things" and "izinto" was

operating in the two versions. It is probable that the word *izinto* "things", used in a general sense was rendered more specific (objects) in the translation into Zulu. Every language has idiomatic expressions which often do not translate very easily and particular care needs to be taken with the translation of such language. However, given that item 35 has only produced a negative item-total correlation in the English version and not in the Zulu version, the problem posed by this item is somewhat less serious. It is after all, the Zulu version that is been targeted for use in this country.

9.2.3 Item-item correlations

If an item in Zulu is equivalent to the same item in English then the responses given by the entire sample to that item in each language, should correlate significantly. In this way, if an instrument in two languages is shown to contain a high level of item-item correlations, the two language versions can be considered to be comparable.

The two items for which English-Zulu item-item correlations did not reach significance were, item 43 - "Have you recently been finding things hard?" and item 54 - "Have you recently been feeling reasonably happy, overall?". It is not surprising that the correlation between the two language versions would be poor for item 43. Item 43 has already been identified as one of the items with a relatively high endorsement discrepancy (see discussion earlier). A possible explanation relating to the use of the word "things" and "izinto" English and Zulu respectively, has been identified and discussed. The endorsement discrepancy between the two languages for item 54 was 0.108. Item 54 therefore just barely missed falling below the cut-off of 0.1 used to separate items of high and low endorsement discrepancies. It is likely that with item 54 too, the phrasing of the question may have evoked different meanings in each of the languages. During the translation process some difficulty was expressed by

the translation teams in finding a Zulu translation for the word
"overall". It is quite likely that this difficulty was not
successfully overcome. Possible rephrasing and further research
would be necessary to resolve the problem with items 43 and 54.

The existence of only two items with non-significant item-item
correlations (Table 7) is an indication of good comparability
between the English and Zulu versions, particularly given the
large number of items with discrepant endorsement frequencies.
The high rate of item-item correlation is an indication of high
fidelity translations. This result is also far better than those
achieved in other attempts at translating the GHQ. Sriram et al.
(1989) found four items with non-significant item-item
correlations between the English and Kannada versions. Eight
items with non-significant item-item correlations were found in
the comparison of the English and Chinese versions of the GHQ
(Chan, 1985). None of the South African studies reviewed here
have reported item-item correlations.

9.3 Reliability coefficients

Both language versions of the GHQ in use showed acceptable levels
of internal consistency on the basis of their respective alpha
coefficients (Table 8). The alpha coefficient of .89 attained for
the Zulu version indicates the relatively high internal
consistency of the scale. This is the highest coefficients
reported for a Zulu version of the GHQ thus far. O'Neil (1988)
reported an alpha of .82 for her Zulu version of the GHQ-30. Leeb
(1986) reported a somewhat lower alpha coefficient of .71 in her
study. The reliability of the English version is by comparison
much lower (.81). One reason for the lower reliability
coefficient of the English version could lie with problematic
phrasing of items in the English version. It has already been
shown that item 35 only produced a negative item-total
correlation in the English version and not in the Zulu version.

The alpha coefficient for the English version of the GHQ-30 improves from .81 to .82 if Item 35 is excluded.

The alpha coefficient of the present study also compares favourably with other translated versions of the GHQ. Resultant alpha coefficients for the Chinese version were .85 (Chan and Chan, 1983) and .93 (Chan, 1985). Sriram et al. reported an alpha of .81 for the Kannada version.

Lower alpha coefficients were obtained for the shorter versions of the GHQ. This is expected, given the smaller number of items involved. Reliability coefficients have not been reported for the shorter forms of the Zulu GHQ used in other studies.

The reliability coefficient for the Zulu version of the GHQ-20 is also at an acceptable level (.84). The same cannot however be said for the Zulu version of the GHQ-12 and some caution is warranted with regard to the use of this Zulu version in the present population. Despite this the coefficient for the Zulu GHQ-12 (.77) is still higher than that achieved by the Zulu GHQ-30 (.71) employed by Leeb (1986).

9.4 Factor analysis

Attempts at factor analysis within studies using the GHQ have been driven largely by the motivation to explore the possible existence of sub-scales within the instrument, even though much of what has previously been said argues for the theoretical and defacto unidimensionality of the GHQ. This is not the motivation underlying the present attempt at factor analysis. For the present purposes, in order to demonstrate equivalence between the English and Zulu versions of the questionnaire, the two versions would need to produce similar factor solutions. If such solutions do emerge, the English and Zulu versions can be said to have attained a state of conceptual equivalence (Sartorius and Kuyken, 1994).

Goldberg (1976) cautions that the "findings of any principal components analysis are dependent upon the nature of the test items and the population on which the test is calibrated, with different items or on a different population, other factors might have been obtained." (p51). Of the studies reviewed, none of the South African studies employing the GHQ attempted factor analysis, let alone factor analysis of the 30-item version with a Zulu-speaking sample. It is most likely that the present study is the first such attempt. While noting Goldberg's words of caution, given the lack of comparable data it is believed that the present discussion can benefit from limited comparisons with other factor analysis studies, albeit that the samples are not similar. For example, it would be both interesting and useful to explore whether the factor solutions of the present study tend to follow notable patterns discovered by Goldberg himself and other workers who have employed the GHQ.

For both versions, several factors emerged with eigenvalues greater than 1; there were ten on the English version and eight on the Zulu version (Table 9). The emergence of several factors with eigenvalues greater than 1 is common for the GHQ. Chan and Chan (1983) also report ten factors on the English version with eigenvalues exceeding unity. The proportions of variance that these two sets of factors (eigenvalues > 1) account for in their respective versions is almost identical, 57.5 on the English version and 57.4 on the Zulu version. The ten factors reported by Chan and Chan (1986) accounted for a somewhat higher percentage (66%) of total variance.

Both the English and Zulu versions approached singled factor solutions. For both versions the first factors accounted for the largest proportion of the variance, namely 17.3% (English) and 25.9% (Zulu) and all subsequent factors accounted for relatively smaller proportions of the variance (< 6.5%). The Zulu version is clearly closer to a one-factor solution with the first factor accounting for more than a quarter of the total variance.

This result of a single-factor solution is not surprising and is in fact a replication of the results obtained by Goldberg (1972) in the early development studies of the GHQ and of the results of other workers. The first factor in Goldberg's study was a large general factor that accounted for 45.6% of the total variance. Goldberg (1972) explains that the emergence of a large general factor should be expected, given that the item reduction techniques employed in selecting items for the GHQ favoured generality rather than specificity. Layton and Rust (1986) also reported a large first general factor, accounting for 23.8% variance in a sample of high school pupils. Likewise, large first factors were also reported by Chan and Chan (1983) and Chan (1985). In this way, the factor solutions of the present study is clearly consistent with general patterns recorded elsewhere.

Given the 'single-factor' nature of both versions it was considered fruitful to compare the first factors in greater detail (Table 10). The larger set of items (23 versus 16) with saturations on Factor 1 of the Zulu version is expected given that Factor 1 one of the Zulu version accounted for 25.9 % of the variance as opposed to the 17.3 % variance accounted for by Factor 1 of the English version.

An examination of the unrotated factor matrixes for Factor 1 (Table 10) shows clear overlap between the English and Zulu versions in terms of the items that loaded heavily (>.40) on each Factor 1. All but one of the sixteen items that loaded heavily on Factor 1 of the English version also loaded heavily on factor one of the Zulu version. In other words, almost 94% of the items on Factor 1 in the English version correspond with items on Factor 1 of the Zulu version. This means that not only do both versions compare well in terms of them both approaching single-factor solutions, but the single factors of each are also very similar in terms of item saturations. The seven items (items 27, 31, 32, 53, 50, 52, 54) which loaded heavily on the Zulu version but did not do so on the English version, all had loadings above .23 on the English version. So, while not making the cut-off of .40,

they nevertheless still had moderate loadings on the English version.

The examination of the matrixes of the six factors after rotation (Table 11) showed the items with high loadings to be factorially simple. No item had high loadings in the same direction on more than one factor. This was the case for both versions. Just two items on the Zulu version (items 30 and 36) loaded on more than one factor, however for both items, the loading was positive on one factor and negative on the other.

In the principal components factoring method used here the unrotated matrix takes out as much variation as possible per factor. The rotated matrix tries to give as clear a pattern as possible. Substantial overlap in terms of item loadings between the first factor on the unrotated matrix and the first factor on the rotated matrix is always encouraging if a unidimensional interpretation of an instrument is sought. Examination of Table 10 and Table 11 shows substantial overlap with regard to the first factor on the unrotated and rotated matrixes.
Such overlap is an indication of the unidimensionality of the instrument.

The six rotated factors on each version have been interpreted and labelled (Table 12). The labels given to factors are based on the constructs implied by the items with heavy loadings on each factor. This exercise of labelling the factors is done merely to aid the process of comparing the two versions of the GHQ and to contrast the present factor solutions against those reported by other workers. As such, the labels are not intended to imply the existence of subscales within the GHQ.

Table 12: Factor labels for the English and Zulu versions

English version

Factor 1 - Anxiety with insomnia

Factor 2 - Depression

Factor 3 - General coping and satisfactory relationships

Factor 4 - Feeling in control

Factor 5 - Social functioning

Factor 6 - Feeling optimistic

Zulu version

Factor 1 - Anxiety with depression

Factor 2 - Depression and difficulty with tasks

Factor 3 - General coping and satisfactory relationships

Factor 4 - Anhedonia

Factor 5 - Insomnia

Factor 6 - Social functioning

The rotated matrixes show fairly similar dimensions underlying the factors on the two versions (Table 12). Common to both versions are dimensions of "anxiety", "depression", "sleep disturbance" on the one hand and dimensions of "general coping and satisfactory relationships" and "social functioning" on the other. The English version produced dimensions of "feeling in control" and "feeling optimistic", both of which were not found on the Zulu version. Likewise, the Zulu version produced a dimension of "anhedonia" which was not found on the English version.

Most of the dimensions found here have been part of factor solutions reported elsewhere. In a four factor solution, Goldberg et al. (1976) reported dimensions of "depression and anxiety", "insomnia and anergia", "social functioning" and "anhedonia" from a sample of white patients; and "depression and anxiety", "anhedonia", "anergia" and "insomnia" from a sample of black

patients. In a five factor solution, Layton and Rust (1986) reported dimensions of "depression and anxiety", "insomnia and anergia", "somatic symptoms", "difficulty coping" and "loss of confidence" from a sample of high school pupils.

In both the Goldberg et al. (1976) as well as the Layton and Rust (1986) studies, a first large factor of "depression and anxiety" emerged. This result was replicated in the present study with the first factor of the Zulu version. In this respect, the English version was different only in that "anxiety" and "depression" fell on separate factors, namely, Factor 1 and Factor 2 respectively. The correspondence between the first factor on the Zulu version and that of the Layton and Rust (1986) study, is particularly significant given that both studies employed samples of high school pupils. Such samples could be considered to comprise more 'normal' individuals relative to the general practice patient sample of Goldberg et al. (1976).

Substantial dimensional overlap occurred on Factor 1. All the items that had saturations on this factor in both versions, had loadings above .45 (Table 11). This direct overlap on Factor 1 between the two versions is based on the 'anxiety' dimension. Four items on the English-Factor 1 refer to "worry", they are; item 14-lost much sleep because of worry?, item 39-been feeling stressed and worried all the time, item 55-been feeling nervous and worried all the time?, item 58-found at times you couldn't do anything because you felt worried?. Three of these items had the highest loadings (.65, .64 and .62) of all the items on the English Factor 1. By contrast, two items on the Zulu-Factor 1 refer to "worry", they are, item 55-been feeling nervous and worried all the time?, 58-found at times you couldn't do anything because you felt worried?. One of these produced the highest loading (.88) on this factor. A significant anxiety dimension is to be expected with the GHQ. It has been hypothesised that " the GHQ is more sensitive to the detection of anxiety symptoms and that anxiety is the primary dimension assessed by the GHQ" (Shek, 1989, p895). Support for this hypothesis can also be found in the

work of Goldberg and Hillier (1979) in establishing the 28-item version. They report a scale consisting of "anxiety and worry" symptoms and note that "anxiety is a core phenomenon which underlies the common syndromes of psychiatric disorder" (p143).

The "anxiety" dimension on the English-Factor 1 was associated with the "insomnia", as contributed by high loadings of items 14 and 20 (.65 and .56 respectively). This combination did not emerge on the Zulu-Factor 1. Although "insomnia" is part of the factor solution of the Zulu version, it fell on an altogether separate factor (Factor 5). Sleep disturbances would appear to be common with student samples (Chan, 1985).

Conversely, the "anxiety" dimension of the Zulu-Factor 1 was associated with "depression". The 'depression' dimension tended to straddle both Factor 1 and Factor 2 on the Zulu version. Depression on the English version fell onto a separate factor (Factor 2). The prominence of anxiety and depression dimensions in factor solutions of the GHQ is well established. Both Chan (1985) and Shek (1989) reported dimensions of anxiety (Factor 1) and of depression (Factor 2) in their factor solutions. Likewise, Chan and Chan (1983) observed dimensions of anxiety (Factor 1) a depression (Factor 3) in their study.

It has been noted that factor solutions are dependent on the sample providing the data. The present sample were high school students in their final and penultimate years at school. These are two important years in the lives of pupils, possibly the most stressful. It is possible that the strong "worry" element in the first factors of this data set, is related to concerns about entering matric for the standard nine pupils in the sample and about matriculating for the rest. While further research would be necessary to test this hypothesis, support for this hypothesis can be taken from the convergence, in terms of anxiety and depression, of the present factor solutions with those of Chan and Chan (1983) and Chan (1985). Both the latter studies employed samples of first-year undergraduate Chinese students. Such

samples would be more similar to that of the present study, relative to the patient samples of other studies, for example, Goldberg et al., 1976 and Medina-Mora et al., 1983.

Factor 3 emerged with the best direct dimension-factor overlap between the factor solutions of the English and Zulu versions. This occurred with the dimension called "social functioning and general coping". The Zulu-Factor 3 contained slightly more (2) items than the English-Factor-3. However, all but one of the items (item 32) on the English-Factor 3 also appear on the Zulu-Factor 3.

Factor 5 (Social functioning) on the English version finds its equivalent in Factor 6 on the Zulu version.

The incongruence between the two versions is made up by Factor 4 and Factor 6 on the English version and Factor 4 on the Zulu version. What is most important about this mismatch, is that the factors on the English version represent what one may call 'healthy functioning' whereas the factor on the Zulu version represents 'dysfunction'. The unmatched factors therefore cast the Zulu version closer towards the 'dysfunctional' end of a hypothetical functional-dysfunctional continuum and the English version closer to the 'functional' or healthy end of the continuum. It was noted that respondents in general scored higher on the Zulu version of the GHQ than they did on the English version. High scores on the GHQ would signal dysfunction. A possible explanation of a social desirability effect has already been posited towards accounting for the difference in scores on the two versions. It is likely that this effect has also influenced the factor solutions in that respondents attempted to present a more healthy/functional image on the English version. This would be the type of impression management that Tyson et al. (1988) referred to in describing the responses of bilinguals.

On the whole, there is substantial correspondence between the factor solutions of the Zulu and English versions of the GHQ. Beyond this, the factors solutions of both versions conform to notable patterns established in other studies. All of this can be interpreted to mean that a level of conceptual equivalence exists between the English and Zulu versions. There exists a commonness with regard to the psychological constructs that underlie the instrument in its two language versions. Given that the factor solutions, particularly that of the Zulu version, follow established trends recorded in other studies, one is offered a level of confidence that the use of the GHQ within a Zulu-speaking population is not illegitimate. The literature on the GHQ reviewed earlier showed indications that psychological constructs underlying the GHQ tend towards universalism. Sartorius and Kuyken (1994) warn against ethnocentric bias in the use of instruments of foreign origin. The present results tend to counter the possibility of claims of ethnocentric bias in employing the GHQ with a Zulu-speaking population.

It is worth emphasising that the examination of the factor structures of the GHQ here was mostly an exploration of the overlap between the English and Zulu versions and the overlap with factor structures recorded elsewhere in the GHQ literature. Limited weight should therefore be attached to the present findings for purposes of practical applications and interpretations. Such caution is offered because extra 'scales' based only on a few items will have a low reliability.

Chapter 10

Conclusion

South Africa is poised at a crucial period in its history and development. Restructuring is taking place at many levels within institutions and society as a whole. Health care is being transformed from a tertiary care system to a primary care system. There are strong calls and compelling reasons for mental health care to also be delivered within primary care facilities. Mental health screening and epidemiological research are of paramount importance to an effective primary health care system. There is consequently a dire need for baseline mental health data for the population as a whole, and particularly for the much neglected African sub-population. Given the shortage of resources this need can only effectively be met via the use of psychometrically sound instruments of proven reliability and validity.

It has been shown that original test development in the country has been inadequate and unrepresentative. There presently exists a substantial reliance on instruments of foreign origin. In all likelihood such reliance will continue in the medium term. A case has been made that very often, such foreign instruments are applied without their applicability to the local testing environment having been tested.

It is thus recommended that instruments of foreign origin should pass through an intensive and extensive research programme so that they may be utilised with confidence and without the trappings of ethnocentric bias. Rigorous multi-stage translation procedures which are evaluated need to be the basis of such research programmes. Once a translation of high fidelity has been acquired, the reliability and validity indices of the instrument need to be established in the local setting. Such a process will allow for norming exercises to proceed. Ultimately, such a research programme will provide the instrument with a solid

foundation that will allow for wider practical utilisation of the instrument. The utility of the instrument is thus greatly advanced because of the confidence with which researchers and other mental health personnel will be able to employ the instrument. An instrument with such a status can more easily contribute to theory building.

The General Health Questionnaire in its original English format and setting has been found to be an effective screening device with patient and community populations. It has repeatedly been shown to be a reliable and valid measure of non-psychotic illness. The instrument is now embedded in a large body of literature. More than forty GHQ studies were reviewed for the present study alone. Likewise, sound psychometric properties have been established with the GHQ in many countries around the world, including a number of countries from the non-western developing part of the world. The instrument has been translated into more than thirty six languages and shows properties of cross-cultural applicability.

In South Africa, the GHQ has been implemented in a few studies. However, such limited use has already seen the employment of a few translated Zulu versions of the instrument. The South African studies have all been investigations into important aspects of mental health and in a sense have been socially responsive. Given the dearth of original local instruments and the widespread usage of the GHQ, it is highly likely that there would be continued use of the instrument in this country. Given the demographic composition of this country, the GHQ will continue to be employed with people who mainly speak an African language. It is therefore necessary and important that the instrument be properly translated into Zulu and the translation be systematically assessed. Until the present study, the translation and psychometric properties of the GHQ have remained untested locally. The present study was therefore seen as germane, essential and original within the South African context.

On the basis of the present results, the GHQ has demonstrated
adequate psychometric characteristics in both the English version
and the translated Zulu version. It has good internal
consistency, the alpha coefficients being similar to those
reported in the literature. The alpha coefficient for the present
Zulu version is the highest yet recorded.

Version equivalence at the scale level is demonstrated by the
high positive-correlations between the two language versions,
including the substantial concordance rate between high-scorers.

Item analysis of the GHQ, however, reveals certain discrepancies
between the two versions as evident from the nine items with
discrepant endorsement frequencies. Other studies have also
reported such discrepancies at the item level with comparisons of
the English version and translated versions (Sriram et al., 1989;
Chan, 1985). While such results may require further work on
particular items, it also attests to the complexity of the
translation process. Sriram et al., argue that it may at times be
simply impossible to obtain precise semantic equivalence to
certain words and phrases in a different language.

Despite the discrepancies on some items with regard to
endorsement frequencies, the item-total correlations and item-
item correlations were overwhelmingly positive and significant.
Only a single item on the English version correlated negatively
with the total and all the items on the Zulu version possessed
positive item-total correlations. Only two items did not achieve
significant item-item correlations.

Perhaps one of the most encouraging findings of this study is the
relatively marked similarity between the English and Zulu factor
structures of the GHQ, and the similarity of such factor
solutions to those reported in other studies. The factor
solutions point to possible areas of application for the GHQ
within a Zulu population. The Zulu version of the GHQ shows signs
of being an effective instrument for mental health screening,

particularly in situations where one wants to detect anxiety and
depression.

In most respects the Zulu version produced in this study has been
shown to be equivalent to the English version. The satisfactory
results of the present study also demonstrates the value of a
rigorous multi-stage translation process for instruments of
foreign origin.

A research programme for the GHQ has been proposed on page 14.
The present study has provided a significant foundation for the
implementation of such a programme. A basis has been laid for
further validity studies which compare the GHQ's case-detecting
properties against that of a standardised clinical interview. In
addition studies that establish norms with various 'normal' and
'patient' samples would need to follow. Such studies will allow
for appropriate cutting-scores to be determined for South African
samples. The present results and those of other South African
studies tend to indicate that the original cutting-scores
recommended by Goldberg may be too low. In this way it is hoped
that the present study has contributed towards the development of
a useful mental health screening instrument.

The government plans to establish several health clinics as part
of the primary health care system. The majority of people
attending such clinics will be from the African population. In
KwaZulu Natal such attendees will be predominately Zulu speaking.
It is believed that the present Zulu version of the GHQ could be
of significant value within such a setting. Although the GHQ has
been employed before, the present Zulu version is more likely to
allow for greater confidence in its use given the data supporting
it. Wider application and ultimately the construction of theory
would thus be enhanced by the efforts of the present study.

References

Aderibigbe, Y.A. & Gureje, O. (1992). The validity of the 28-item General Health Questionnaire in a Nigerian antenatal clinic. <u>Social Psychiatry and Psychiatric Epidemiology</u>, <u>27</u>, 280-283.

Aitchison, J.J.W. (1995). <u>The demography of functional illiteracy among adults in South Africa</u>. Pietermaritzburg: Centre for Adult Education, University of Natal.

American Association for the Advancement of Science (AAAS) (1990). <u>Apartheid Medicine: health and human rights in South Africa</u>. Washington: American Association for the Advancement of Science.

Anastasi, A. (1982). <u>Psychological Testing</u>, 5th edition, New York: MacMillan.

Banks, M.H. (1983). Validation of the General Health Questionnaire in a young community sample. <u>Psychological Medicine</u>, <u>13</u>, 349-354.

Banks, M.H. & Jackson, P.R. (1982). Unemployment and risk of minor psychiatric disorder in young people: cross-sectional and longitudinal evidence. <u>Psychological Medicine</u>, <u>12</u>, 789-798.

Banks, M.H., Clegg, C.W., Jackson, P.R., Kemp, N.J., Stafford, E.M. & Wall, D.T. (1980). The use of the General Health Questionnaire as an indicator of mental health in occupational studies. <u>Journal of Occupational Psychology</u>, <u>53</u>, 187-194.

Bluen, S.D. & Odesnik, J. (1988). Township unrest: Development of the Township Life Events Scale. <u>South African Journal of Psychology</u>, <u>18</u>, 50-57.

Bond, M.H. & Yang, K. (1982). Ethnic affirmation versus cross-cultural accommodation: the variable impact of questionnaire language on Chinese bilinguals from Hong Kong. Journal of Cross-Cultural Psychology, 13, 169-185.

Brand, H.J., van Noordwyk, J.S.J & Hanekom, J.D.M. (1994). Die toepassing van die Self-Directed Search op 'n groep swart adolessente. South African Journal of Psychology, 24, 47-52.

Brenner, S.-O. & Bartell, R. (1983). The psychological impact of unemployment: a structural analysis of cross-sectional data. Journal of Occupational Psychology, 56, 129-136.

Brislin, R. (1970). Back-translation for cross-cultural research. Journal of Cross-Cultural Psychology, 1, 185-216.

Brislin, R.W. (1980). Translation and Content Analysis of Oral and Written Materials. In H.C. Triandis & J.W. Berry (eds). Handbook of Cross-Cultural Psychology: Methodology. Vol 2. London: Allyn and Bacon.

Brislin, R.W. (1986). The wording and translation of research instruments. In W.J. Lonner & J.W. Berry (eds). Field methods in Cross-Cultural Research. Beverly Hills: Sage.

Broomhall, H.S. & Winefield, A.H. (1990). A comparison of the affective well-being of young and middle-aged unemployed men matched for length of unemployment. British Journal of Medical Psychology, 63, 43-52.

Burvill P.W. & Knuiman, M.W. (1983). The influence of minor psychiatric morbidity on consulting rates to general practitioners. Psychological Medicine, 13, 635-643.

Cairns, E., Wilson, R., McClelland, R. & Gillespie, K. (1989). Improving the validity of the GHQ30 by rescoring for chronicity: a failure to replicate. Journal of Clinical Psychology, 45, 793-798.

Candell, G.L. & Hulin, C.L. (1986). Cross-language and cross-cultural comparisons in scale translation - independent sources of information about item nonequivalence. Journal of Cross-Cultural Psychology, 17(4), 417-440.

Carlyle, R. A. (1991). Political violence, displacement and township families: an exploratory study. Unpublished master's thesis, University of Natal, Pietermaritzburg.

Central Statistical Services (1995). October Household Survey 1994. Pretoria: Central Statistical Services

Chan, D.W. (1985). The Chinese version of the General Health Questionnaire: does language make a difference?. Psychological Medicine, 15, 147-155.

Chan, D.W. & Chan, T.S.C. (1983). Reliability, validity and the structure of the General Health Questionnaire in a Chinese context. Psychological Medicine, 13, 363-371.

Church, A.T., Katigbak, M.S. & Castaneda, I. (1988). The Effects of language of data collection on derived conceptions of healthy personality with Filipino bilinguals. Journal of Cross-Cultural Psychology, 19, 178 - 192.

Dawes, A. (1994). The emotional impact of political violence.. In A. Dawes & D. Donald (eds). Childhood and Adversity: psychological perspectives from South African research. Cape Town: David Philip.

Dhadphale, R., Ellison, R.H. & Griffin, L. (1983). The frequency of psychiatric disorders among patients attending semi-urban and rural general out-patient clinics in Kenya. British Journal of Psychiatry, 142, 379-383.

Duckitt, J. (1993). Further validation of a subtle racism scale for South Africa. South African Journal of Psychology, 23, 116-119.

Duckitt, J. (1991). The development and validation of a subtle racism scale in South Africa. South African Journal of Psychology, 21, 233-239.

Dunnigan, T., McNall, M. & Mortimer, J. T. (1993). The problem of metaphorical nonequivalence in cross-cultural survey research: comparing the mental health statuses of Hmong refugee and general population adolescents. Journal of Cross-Cultural Psychology, 24, 344-365.

Ellis, B. B., Becker, P. & Kimmel, H. D. (1993). An Item Response Theory evaluation of an English version of the Trier Personality Inventory (TPI). Journal of Cross-Cultural Psychology, 24, 133-148.

Ellis, B.B. & Kimmel, H.D. (1992). Identification of Unique Cultural Response Patterns by means of Item Response Theory. Journal of Applied Psychology, 77, 177 - 184.

Finlay-Jones, R.A. & Murphy, E. (1979). Severity of Psychiatric Disorder and the 30-item General Health Questionnaire. British Journal of Psychiatry, 134, 609-616.

Fontanesi, F., Gobetti, C., Zimmerman-Tansella, C.H. & Tansella, M. (1985). Validation of the Italian version of the GHQ in general practice. Psychological Medicine, 15, 411-415.

Gerdes, L. C. (1992). Impressions and questions about psychology and psychologists. South African Journal of Psychology, 22, 39-43.

Goldberg, D & Huxley, P. (1980). Mental illness in the community: the pathway to psychiatric care. London: Tavistock.

Goldberg, D. P. (1972). The detection of psychiatric illness by questionnaire. London: Oxford University Press.

Goldberg, D.P. & Blackwell, B. (1970). Psychiatric illness in general practice: a detailed study using a new method of case identification. British Medical Journal, 2, 439-443.

Goldberg, D.P. & Hillier, V.F. (1979). A scaled version of the General Health Questionnaire. Psychological Medicine, 9, 139-145.

Goldberg, D.P., Rickels, K., Downing, R. & Hesbacher, P. (1976). A comparison of two psychiatric screening tests. British Journal of Psychiatry, 129, 61-67.

Goodchild, M.E. & Duncan-Jones, P. (1985). Chronicity and the General Health Questionnaire. British Journal of Psychiatry, 146, 55 - 61.

Govender, K. (1995). An investigation of the role of perceived sources of stress, perception of work environment, type of hospital ward and nurse rank in occupational distress, coping and burnout among practising nurses. Unpublished master's thesis. University of Natal, Pietermaritzburg.

Gureje, O. & Obikoya, B. (1990). The GHQ-12 as a screening tool in a primary care setting. Social Psychiatry and Psychiatric Epidemiology, 25, 276-280.

Gurney, R. M. (1980). The effects of unemployment on the psycho-social development of school-leavers. <u>Journal of Occupational Psychology</u>, <u>53</u>, 205-213.

Hambleton, R. K., Swaminathan, H. & Rogers, H.J. (1991). <u>Fundamentals of Item Response Theory</u>. London: Sage.

Hepworth, S.J. (1980). Moderating factors of the psychological impact of unemployment. <u>Journal of Occupational Psychology</u>, <u>53</u>, 139-146.

Hulin, C.L., & Mayer, L.J. (1986). Linguistic and Psychometric equivalence of items across languages: an evaluation of a translation of the JDI into Hebrew. <u>Journal of Applied Psychology</u>, <u>71</u>, 83-94.

Huysamen, G.K. (1983). <u>Psychological measurement: an introduction with South African examples</u>. Pretoria: Academica.

Irvine, S.H. & Carroll, W.K. (1980). Testing and Assessment across cultures. In H.C. Triandis & J.W. Berry (eds). <u>Handbook of Cross-Cultural Psychology: Methodology</u>. Vol 2. London: Allyn and Bacon.

Jackson, P.R., Stafford, E.M., Banks, M.H. & Warr, P.B. (1983). Unemployment and Psychological Distress in Young People: the moderating role of employment commitment. <u>Journal of Abnormal Psychology</u>, <u>68</u>, 525-535.

Jenkins, R., Macdonald, A., Murray, J. & Strathdee, G. (1982). Minor psychiatric morbidity and the threat of redundancy in a professional group. <u>Psychological Medicine</u>, <u>12</u>, 799-807.

Kemp, N.J. & Mercer, A. (1983). Unemployment, disability and rehabilitation centres and their effects on mental health. <u>Journal of Occupational Psychology</u>, <u>56</u>, 37-48.

Kriegler, S. (1993). Options and directions for psychology within a framework for mental health services in South Africa. South African Journal of Psychology, 23, 64-70.

Layton, C. & Rust, J. (1986). The factor structure of the 60 item General Health Questionnaire. Social Behavior and Personality Research, 14, 123-131.

Leeb, W. (1986). Use of the General Health Questionnaire to evaluate the psychological effects of unemployment. Unpublished honour's study. University of Natal, Pietermaritzburg.

Liddell, C., Kvalsvig, J., Qotyana, P. & Shabalala, A. [1992]. Community violence and levels of aggression in young South African children. Unpublished manuscript. Human Sciences Research Council, Pretoria.

Lindegger, G. & Ward, G. (1995). The AIDS crisis: review of psychological issues and implications, with special reference to the South African situation. South African Journal of Psychology, 25, 1-11.

Lobo, A., Perez-Echeverria, M.J., Artal, J. (1986). Validity of the scaled version of the General Health Questionnaire (GHQ-28) in a Spanish population. Psychological Medicine, 16, 135-140.

Lonner, W. J. (1990). An overview of cross-cultural testing and assessment. In R.W. Brislin (ed). Applied cross-cultural psychology. London: Sage.

Lonner, W.J. (1980). The Search for Psychological Universals. In H.C. Triandis & W.W. Lambert (eds). Handbook of Cross-Cultural Psychology: Perspectives. Vol. 1. Boston: Allyn and Bacon.

Mari, J.J. & Williams, P. (1984). Minor psychiatric disorder in primary care in Brazil: pilot study. Psychological Medicine, 14, 223-227.

Mari, J.J. & Williams, P. (1985). A comparison of the validity of two psychiatric screening questionnaires (GHQ-12 and SRQ-20) in Brazil using relative operating characteristics (ROC) analysis. Psychological Medicine, 15, 651-659.

Marin, G., Betancourt, H. & Kashima, Y. (1983). Ethnic affirmation versus social desirability: explaining discrepancies in bilinguals' responses to a questionnaire. Journal of Cross-Cultural Psychology, 14, 173-186.

McGarr, C.J. (1985). The influence of race on the perceived leadership role of the first-line supervisor. Unpublished master's thesis. University of Natal, Pietermaritzburg.

Medina-Mora, M.E., Padilla, G.P., Campillo-Serrano, C., Mas, C.C., Ezban, M., Caraveo, J. & Corona, J. (1983). The factor structure of the GHQ: a scaled version for a hospital's general practice service in Mexico. Psychological Medicine, 13, 355-361.

Meehl, P.E. (1978). Theoretical risks and tabular: Sir Karl, Sir Ronald, and the slow progress of soft psychology. Journal of Consulting and Clinical Psychology, 46, 806-834.

Michelson, C.L. (1991). Trauma, levels of distress, and coping amongst displacees in the Pietermaritzburg area. Unpublished master's thesis. University of Natal, Pietermaritzburg.

Miller, P. McC. & Surtees, P.G. (1991). Psychological Symptoms and their Course in First-Year Medical Students as assessed by the Interval General Health Questionnaire (I-GHQ). British Journal of Psychiatry, 159, 199-207.

Miller, T. & Swartz, L. (1992). Psychology and epidemiology: an uncomfortable alliance?.South African Journal of Psychology, 22, 52-58.

Minsel, B., Becker, P. & Korchin, S. J. (1991). A cross-cultural view of positive mental health: Two orthogonal main factors replicable in four countries. Journal of Cross-Cultural Psychology, 22, 157-181.

Nell, V. (1994). Interpretation and misinterpretation of the South African Wechsler-Bellevue Adult Intelligence Scale: a history and a prospectus. South African Journal of Psychology, 24, 100-109.

Nell, V. (1990). One world, one psychology: 'Relevance' and ethnopsychology. South African Journal of Psychology, 20, 129-140.

O'Neil, D.F. (1988). The role of the shop steward in South African industrial relations. Unpublished doctoral thesis. University of Natal, Pietermaritzburg.

Ormel, J., Koeter, M.W.J., & van den Brink, W. (1989). Measuring change with the General Health Questionnaire (GHQ): the problem of retest effects. Social Psychiatry and Psychiatric Epidemiology, 24, 227-232.

Owen, K (1991). Test bias: the validity of the Junior Aptitude Test (JAT) for various population groups in South Africa regarding constructs measured. South African Journal of Psychology, 21, 112-118.

Pareek, U & Rao, T.V. (1980). Cross-cultural surveys and interviewing. In H.C. Triandis & J.W. Berry (eds). Handbook of Cross-Cultural Psychology: Methodology. Vol 2. London: Allyn and Bacon.

Parsons, C. K. & Hulin, C. L. (1982). An empirical comparison of Item Response Theory and Hierarchical Factor Analysis in applications to the measurement of Job Satisfaction. Journal of Applied Psychology, 67, 826-834.

Perkel, A. K. (1988). Perceived locus of control as a correlate of traumatisation during detention: with implications for therapeutic intervention. Unpublished master's thesis. University of Natal, Pietermaritzburg.

Pillay, A.L. (1987). Models of consultation in the practice of paediatric psychology. Psychotherapeia, 14, 19-23.

Pillay, Y. (1995). Personal correspondence. University of Durban-Westville, Durban.

Pillay, Y., Magwaza, A. & Petersen, I. (1992). Civil conflict in Mpumulanga: some mental health sequelae in a sample of children and their primary caregivers. South African Journal of Child and Adolescent Psychiatry, 4(2), 42-45.

Poortinga, Y.H. (1992). Towards a conceptualization of culture for psychology. In S. Iwawaki, Y. Kashima & K. Leung (eds). Innovations in Cross-Cultural Psychology. Amsterdam: Swets and Zeitlinger.

Rack, P. (1982). Race, Culture and Mental Disorder. London: Tavistock.

Rawlins, R.C.A. (1983). Attitudes to mental retardation: a cross cultural study. Unpublished master's thesis. University of Natal, Pietermaritzburg.

Retief, A. (1988). Method and theory in cross-cultural psychological assessment. Pretoria: Human Sciences Research Council.

Sang, D. (1992). Research Questionnaire and Checklists being used with Indochinese refugees: a survey. In S. Iwawaki, Y. Kashima & K. Leung (eds). Innovations in Cross-Cultural Psychology. Amsterdam: Swets and Zeitlinger.

Sartorius, N. & Kuyken, W. [1994]. Translation of health status instruments. In J. Orley & W. Kuyken (eds). Quality of life assessment in health care settings. Berlin: Springer-Verlag.

Schweitzer, R.D. (1977). Categories of experience amongst the Xhosa: a psychological study. Unpublished manuscript, Rhodes University, Grahamstown.

Sechrest, L., Fay, T.L. & Zaidi, S.M.H. (1972). Problems of translation in cross-cultural research. Journal of Cross-Cultural Psychology, 3, 41-56.

Seedat, M. & Nell, V. (1992). Authoritarianism and autonomy: conflicting value systems in the introduction of psychological services in a South African primary health care system. South African Journal of Psychology, 22, 185-193.

Sharratt, P.A. & van der Heuvel, E (1995). Metamemorial knowledge in a group of black South African school children. South African Journal of Psychology, 25, 59-73.

Shek, D.T.L. (1989). Validity of the Chinese version of the General Health Questionnaire. Journal of Clinical Psychology, 45, 890-897.

Skuse, D, & Williams, P. (1984). Screening for psychiatric disorder in general practice. Psychological Medicine, 14, 365-377.

Sriram, T.G., Chandrashekar, C.R., Isaac, M.K. & Shanmugham, V. (1989). The General Health Questionnaire (GHQ): Comparison of the English version and a translated Indian version. Social Psychiatry and Psychiatric Epidemiology, 24, 317-320.

Stafford, E.M., Jackson, P.R. & Banks, M.H. (1980). Employment, work involvement and mental health in less qualified young people. Journal of Occupational Psychology, 53, 291-304.

Stanley, B. & Gibson, A.J. (1985). The prevalence of chronic psychiatric morbidity: a community sample. British Journal of Psychiatry, 146, 372-376.

Subedar, H. (1993). Screening for mental disorders in the Mahlabatini ward in the Ngcolosi district of the Valley of a Thousand Hills. Unpublished master's thesis. University of Natal, Durban.

Surtees, P.G. & Miller, P. McC. (1990). The Interval General Health Questionnaire. British Journal of Psychiatry, 157, 679-686.

Surtees, P.G. & Tansella, M. (1990). The Interval General Health Questionnaire and its relationship with psychiatric disorder in community and general practice samples. British Journal of Psychiatry, 157, 686-693.

Tarnopolsky, A., Hand, D.J., Mclean, E.K., Roberts, H. & Wiggins, R.D. (1979). Validity and uses of a screening questionnaire (GHQ) in the community. British Journal of Psychiatry, 134, 508-515.

Taylor, C. A., Fourie, R. J. & Koorts, A. S. (1995). Die ontwerp van 'n instrument om verandering in houdings van wit en swart kinders te meet. South African Journal of Psychology, 25, 116-121.

Taylor, T.R. & Boeyens, J.C.A. (1991). The comparability of the scores of blacks and whites on the South African Personality Questionnaire: an exploratory study. South African Journal of Psychology, 21, 1-11.

Triandis, H. C., Vassiliou, V., Vassiliou G., Tanaka, Y. & Shanmugam, A.V. (1972). The analysis of subjective culture. New York: Wiley.

Triandis, H.C. (1980). Introduction to Handbook of Cross-Cultural Psychology. In H.C. Triandis & W.W. Lambert (eds). Handbook of Cross-Cultural Psychology: Perspectives. Vol. 1. Boston: Allyn and Bacon.

Turton, R.W., Straker, G. & Moosa, F. (1991). Experiences of violence in the lives of township youths in 'unrest' and 'normal' conditions. South African Journal of Psychology, 21, 77-84.

Tyson, G.A., Doctor, E.A. & Mentis, M. (1988). A psycholinguistic perspective on bilinguals' discrepant questionnaire responses. Journal of Cross-Cultural Psychology, 19, 413-426.

Vachon, M.L.S., Sheldon, A.R., Lancee, W.J., Lyall, W.A.L., Rogers, J. & Freeman, J.J. (1982). Correlates of enduring distress patterns following bereavement: social network, life situation and personality. Psychological Medicine, 12, 783-788.

van Ede, D.M. (1995). Adapting the Metamemory in Adulthood (MIA) Questionnaire for cross-cultural application in South Africa. South African Journal of Psychology, 25, 74-79.

van Eeden, R. & Visser, D. (1992). The validity of the Senior South African Scale-Revised (SSAIS-R). South African Journal of Psychology, 22, 163-171.

Vasquez-Barquero, J.L., Padierna Acero, I.A., Marton, C.P. & Ochoteco, A. (1985). The psychiatric correlates of coronary pathology: validity of the GHQ-60 as a screening instrument. Psychological Medicine, 15, 589-596.

Viljoen, G., Levett, A, Tredoux, C. & Anderson, S. (1994). Using the Bender Gestalt in South Africa: some normative data for Zulu-speaking children. South African Journal of Psychology, 24, 145-151.

Ward, C. & Sethi, R. (1986). Cross-cultural validation of the Bem Sex Role Inventory: Malaysian and South Indian research. Journal of Cross-Cultural Psychology, 17, 300-314.

Werner, O. & Campbell, D.T. (1970). Translating, working through interpreters, and the problem of decentering. In R. Naroll & R. Cohen (eds). A handbook of method in cultural anthropology. New York: Columbia University Press.

Winefield, A.H. & Tiggemann, M. (1985). Psychological correlates of employment and unemployment: Effects, predisposing factors, and sex differences. Journal of Occupational Psychology, 58, 229-242.

Yang, K. & Bond, M.H. (1980). Ethnic affirmation by Chinese bilinguals. Journal of Cross-Cultural Psychology, 11, 411-425.

APPENDIX A

Some guidelines for the translation of the GHQ into Zulu

Your primary task is to translate each item of the GHQ into Zulu, such that the meaning of the item is retained while the language you use is understandable to a wide range of Zulu speaking people - urban and rural, modern and traditional.

A general rule is to use short simple sentences using the active voice rather than the passive voice. Avoid scholarly terminology. Below, you will find more guidelines on the different forms of equivalence that would have to be achieved in translating the GHQ to Zulu.

a) *Vocabulary equivalence*

For English terms for which there are no Zulu equivalents, use a short description in Zulu that would convey the meaning of the term. Try as far as possible to keep each sentence to less than sixteen words.

b) *Idiomatic equivalence*

The GHQ contains a few idiomatic expressions. These may be replaced by popular Zulu idioms if there are such or by a phrase that conveys the meaning of the idiom.

Zulu idioms may be used in translating some of the other items provided that such idioms are deemed to best convey the meaning and that such Zulu idioms are widely used expressions.

c) *Grammatical-syntactical equivalence*

It is difficult to specify rules to guide one with this problem. The general rule of comprehensibility should serve as a guideline. Avoid convoluted constructions that aid grammatical considerations at the expense of meaning.

d) *Experiential equivalence*

In order for the translation to Zulu to be successful make use of terms that refer to real things and experiences which are likely to be culturally familiar to a wide range of Zulu-speaking people.

e) *Conceptual equivalence*

Ensure that the concepts referred to in the English version are culturally familiar to Zulu-speaking people and that when translated, they imply the same things in Zulu.

d) *Context & Redundancy*

Provide ample context for difficult terms or concepts. Sentences should be constructed such that they aid the understanding of difficult terms or concepts included in them. Redundancy, saying the same things in different ways, may also be used to aid the understanding of particularly difficult terms or concepts.

APPENDIX B

Goldberg's original 30-item version of the General Health Questionnaire.

Have you recently:

7. been able to concentrate on whatever you're doing? (ss)

14. lost much sleep over worry? (uu 64, 66)

20. been having restless, disturbed nights? 66

21. been managing to keep yourself busy and occupied?

26. been getting out of the house as much as usual?

27. been managing as well as most people would in your shoes?

28. felt on the whole you were doing things well?

30. been satisfied with the way you've carried out your task?

31. been able to feel warmth and affection for those near to you?

32. been finding it easy to get on with other people?

33. spent much time chatting with people?

35. felt that you are playing a useful part in things?

36. felt capable of making decisions about things?

39. felt constantly under strain?

40. felt you couldn't overcome your difficulties?

41. been finding life a struggle all the time?

42. been able to enjoy your normal day-to-day activities?

43. been taking things hard?

45. been getting scared or panicky for no good reason?

46. been able to face up to your problems?

47. found everything getting on top of you?

49. been feeling unhappy and depressed?

50. been losing confidence in yourself?

51. been thinking of yourself as a worthless person?

52. felt that life is entirely hopeless?

53. been feeling hopeful about your own future?

54. been feeling reasonably happy, all things considered?

55. been feeling nervous and strung-up all the time?

56. felt that life isn't worth living?

58. found at times you couldn't do anything because your nerves were too bad?

APPENDIX C

The English version of the GHQ-30 administered to the bilingual student sample

Note: Item numbers were excluded from the form given to students.

Ref. No._____

Todays date is _____

Std: _____   Age:_____   Sex: Male___   Female___ (Tick)

Home language: Zulu ___   English ___ Other _____ (Write in)

Instructions

The aim of this exercise is not to examine your health. Your answers will be used to see whether these questions are understandable and whether they work well.

Please answer each question with Yes or No. (✓ Tick your answer.)

Please remember that questions refer to the immediate past or the recent period of your life. When answering each question, think about the recent past.

_____

Have you recently:

7. been able to pay attention to whatever you're doing?	Yes __	No __
14. lost much sleep because of worry?	Yes __	No __
35. felt that you are playing a useful part in things?	Yes __	No __
36. felt able to make decisions about things?	Yes __	No __
39. been feeling stressed and worried all the time	Yes __	No __
40. felt you could not solve your problems?	Yes __	No __
42. been able to enjoy your normal daily activities?	Yes __	No __
46. been able to face your problems?	Yes __	No __
49. been feeling unhappy and depressed?	Yes __	No __
50. been feeling unsure of yourself?	Yes __	No __
51. been thinking of yourself as a useless person?	Yes __	No __
54. been feeling reasonably happy, overall?	Yes __	No __
21. been able to keep yourself busy and occupied?	Yes __	No __
26. been getting out of the house as much as usual?	Yes __	No __
28. felt you were doing most things well?	Yes __	No __

Have you recently:

30. been happy with the way you have done your work?   Yes __ No __

43. been finding things hard?   Yes __ No __

47. found everything going against you?   Yes __ No __

55. been feeling nervous and worried all the time?   Yes __ No __

58. found at times you couldn't do anything because you felt worried?   Yes __ No __

20. been having restless, disturbed nights?   Yes __ No __

27. been managing as well as most people would in your situation?   Yes __ No __

31. been able to feel love and affection for those close to you?   Yes __ No __

32. been able to get on with other people?   Yes __ No __

33. spent much time talking with people?   Yes __ No __

41. been finding life is a problem all the time?   Yes __ No __

45. been getting scared or nervous for no good reason?   Yes __ No __

52. felt that life is mostly hopeless?   Yes __ No __

53. been feeling hopeful about your own future?   Yes __ No __

56. felt that life is not worth living?   Yes __ No __

Thank you for answering these questions.

Please check that you have made a tick for all 30 questions.

Your Name: _____

Your Surname: _____

APPENDIX D

The Zulu version of the GHQ-30 administered to the bilingual student sample

Note: Item numbers were excluded from the form given to students.

Ref. No. _____

Usuku lwanamhlanje _____

Ibanga: _____   Iminyaka:_____

Ubulili: Isilisa____ Isimame____ (✓ Uphawu)

Ulwimi lwaasekhaya: Isizulu ___ Isingisi ___ Olunye _____(Gcwalisa)

Imithetho

Inhloso yami akuyona ukuhlola impilo yenu. Ngizosebenzisa izimpendulo zenu ukuhlola ukuthi lemibuzo isebenza kahle yini.

Ngicela uphendule umbuzo ngamunye ngo **Yebo** noma **Cha**. (✓ Uphawu)

Ngicela ukhumbule ukuthi lemibuzo ikhuluma ngento esanda kwenzeka noma esanda kudlula empilweni yakho.
Uma uphendula umbuzo, khumbula ngezinto esandakudlula.

Esikhathini esimaduze nje, ngabe:

7. ubukwazi ukugxila ngokugcwele kunoma yini oyenzayo?   Yebo ___ Cha ___

14. bulahlekelwa ubuthongo ngokuba nezingxaki?   Yebo ___ Cha ___

35. ubuzwa sengathi izinto ozenzayo zibalulekile?   Yebo ___ Cha ___

36. ubuzizwa ukwazi ukuthatha izinqumo ngezinto?   Yebo ___ Cha ___

39. ubuzizwa uhlala njalo usebunzimeni?   Yebo ___ Cha ___

40. ubuzizwa ungakwazi ukuqeda ubunzima obukukhungethe?   Yebo ___ Cha ___

42. ubukwazi ukuthakasela izinto ojwayele ukuzenza nsukuzonke?   Yebo ___ Cha ___

46. ubukwazi ukubhekana umele izinkinga zakho?   Yebo ___ Cha ___

49. ubuzizwa ungeneme futhi umoya wakho ushone phansi?   Yebo ___ Cha ___

50. uke walahlekelwa ukuzethemba?   Yebo ___ Cha ___

51. uke wazicabanga ungumuntu ongelutho?   Yebo ___ Cha ___

54. ubuzizwa unokwenama okukahle nje, zonke izinto zihlanganisiwe?
Yebo ___ Cha ___

21. ubukwazi ukuzigcina umatasa futhi kuhlale kukhona okwenzayo?   Yebo ___ Cha ___

26. ubuphuma endlini uhambe izikhathi eziningi njengokujwayelekile?   Yebo ___ Cha ___

Esikhathini esimaduze nje, ngabe:

28. ubuzizwa sengathi zonke izinto usazenza kahle?   Yebo ___ Cha ___

30. ubugculisekile ngendlela umsebenzi wakho owenze ngayo?   Yebo ___ Cha ___

43. ubuthatha noma wamukele kanzima izinto ezenzekayo?   Yebo ___ Cha ___

47. uthole zonke izinto zikucindezela?   Yebo ___ Cha ___

55. ubuzizwa unokungakhululeki kahle futhi umpintshekile
sonke isikhathi?   Yebo ___ Cha ___

58. ubuthola kwezinye izikhathi ungakwazi kwenza lutho ngenxa
yokushaywa uvalo?   Yebo ___ Cha ___

20. bekungafiki ubuthongo, kunokuphazamiseka ebusukwini bakho?   Yebo ___ Cha ___

27. ubusaqhuba kahle njengoba bonke abantu bengaqhuba uma
besesimweni esifana nesakho?   Yebo ___ Cha ___

31. uke wakuthola ukufudumala nothando kulabo osondelene nabo?   Yebo ___ Cha ___

32. ubusakhona ukuphilelana nabanye abantu?   Yebo ___ Cha ___

33. uke wasichitha isikhati uxoxisana nabantu?   Yebo ___ Cha ___

41. ubuthola impilo ingumzabalazo sonke isikhathi?   Yebo ___ Cha ___

45. ubuzithola unokwesaba noma ukutatazela ngaphandle kwesizathu?   Yebo ___ Cha ___

52. uke waphelelwa nya yithemba empilweni?   Yebo ___ Cha ___

53. ubuzizwa unethemba ngekusasa lakho?   Yebo ___ Cha ___

56. ubuzwa ukuthi akunasidingo sokuphila?   Yebo ___ Cha ___

Ngiyabonga ngokuthi uphendule lemibuzo

Ngicela uhlole ukuthi lemibuzo engu 30 uyithikhe yonke yini.

Igama lakho _____

Isibongo sakho _____