

MIGRATION ANALYSIS IN SOUTH AFRICA

Utilising the opportunities and strengths; reducing the threats and weaknesses

HSRC RESEARCH OUTPUTS

3322

by Pieter Kok (HSRC)
(E-mail: pkok@hsrc.ac.za)

Paper presented at the
Joint Population Conference
Durban, 4-8 Oct 2004

This paper discusses the potential utility of some of the available data sets for modelling and other national and local migration analyses, and emphasises the need for and potential of longitudinal migration research. A key objective of the paper is to elaborate on the potential utility of theory-based migration modelling. The emphasis in the paper is thus more on theory-based suggestions for linking the different data sources during modelling than on actual applications of the latest census data. This is done by means of a conceptual framework that incorporates most of the latest theoretical contributions and at the same time also attempts to make provision for the often-opposing paradigmatic viewpoints of both the structuralists and the behaviourists.¹

PROBLEM UNDER STUDY

In South Africa the analysis of migration trends was hampered by the absence of comprehensive and detailed data on internal migration. Until recently migration analysts had to rely on sample surveys that are neither detailed nor comprehensive enough for an understanding of this dynamic phenomenon. Only censuses potentially offer information with the required breadth of spatial detail.² Before 1996 South African censuses failed to record migration data such as the place and timing of migratory moves within and to South Africa.

Only a handful of South African researchers have experience in truly in-depth migration analysis. This is due partly to the historical absence of the necessary contextual (census) data, and also to the historical absence of demographers with sufficiently focused migration-analysis experience to provide the required post-graduate education and hands-on training. Experienced South African migration researchers also tend to have been schooled in other disciplines such as economics, geography, history, planning and sociology, therefore lacking the skills of the trained analytical demographer.³

At the same time it should be acknowledged that migration is probably the most complex of the three basic components of population dynamics. The other two components, fertility and mortality, although intricate in their own right, do not have to deal with the same multi-level (macro, meso and micro) influences and impacts as migration. The central position of the multifarious spatial dimension adds significantly to the complexity of migration analysis.

¹ During the 2003 Demsa Conference I attempted to show how these quantitative linkages of different data sources could be made, using a value-expectancy decision-making framework (P Kok. 2003. *Migration in southern Africa: Theoretical, methodological and policy issues*. Paper delivered at the Joint Population Conference, Potchefstroom, 13-17 October 2003). I also tried to show that migration models using data from all three levels are viable and quite effective in explaining not only individual-level migration intentions but also (1996-based) meso-level aggregated actual migration behaviour. As indicated then, Census 2001 data can hopefully be used in scenario-based simulation and statistical models to also analyse and predict the possible changing patterns and trends that may be occurring over time. However, the (perceived) persistent non-availability of spatially detailed and full-coverage migration data from Census 2001 makes it difficult to be more specific at this stage.

² Like almost all countries exhibiting a similar level of development, South Africa does not maintain a rigorous population register of where people currently reside and moved from. (Unlike in some other countries, the South African register is not accessible for research purposes anyway.)

³ This statement should by no means be construed as some kind of criticism against the migration research conducted by scholars from these disciplines. Quite the contrary: without the important contributions over time by, for example, economists, geographers, historians and sociologists, our collective understanding of migration would have been virtually non-existent, and their research is indispensable for the demography-trained migration analyst. My appeal at this (population) conference is merely that South African universities with schools in Demography or Population Studies should endeavour to produce more migration researchers capable of utilising the important tools of demographic analysis.

OBJECTIVES

A kind of SWOT (strengths, weaknesses, opportunities and threats) analysis around migration analysis in South Africa may be needed here. The *opportunities* offered by recent censuses (by Statistics South Africa) and sample surveys (e.g. by the HSRC) can be used effectively for migration analyses. There is a number, albeit small, of productive migration analysts in South Africa. This can certainly be regarded as a potential *strength*, but the concomitant *threat* is that this is probably almost a "dying breed". Furthermore, many of these migration analysts come from disciplines other than Demography, indicating the potential *weakness* that many of these researchers may not be fully equipped to deal with the complex demands of in-depth migration analysis. These problems can only be addressed if South African universities that offer post-graduate courses in Demography or Population Studies produce well-trained migration analysts. Analytical demographers are needed to properly identify and rectify migration-data problems (for example with age-sex distributions), to conduct migration analyses using appropriate indirect methods (also as part of the data-validation process), and to express considered opinions (especially regarding the demography-related nature and consequences of migration).

Two census data sets (1996 and 2001) and two national sample-survey data sets (namely the 2000 and 2001-02 HSRC Migration Surveys) are therefore used to show how to optimise the opportunities that the available data provide for national migration analyses.⁴ The paper emphasises the benefits of longitudinal studies, and indicates how the HSRC's recent migration surveys can be used as a basis for longitudinal research on the causes of migration and non-migration.⁵ The opportunities offered by qualitative research for obtaining insights into intricate migration processes are also highlighted. Migration theory is still too weak to explain the complex dynamics of spatial mobility in South and southern Africa, and the paper attempts to show how to go about trying to fill some of these conceptual gaps.

BACKGROUND

Census 1996 was a welcome exception to the earlier trend in censuses of ignoring South Africa's migration-information needs, and Census 2001 has since added much potential value by also creating opportunities for some trend analyses.⁶ However, it should be noted that all census data on migration have limitations, and these two censuses are no exceptions. Cost considerations dictate that the number of migration questions that can be covered in censuses is limited. Furthermore, South African migration analysts have to deal with some data constraints that may or may not be common in other parts of the world. Not only is the absence of historical data, referred to earlier, a major constraint, but we also have complex definition issues (especially concerning urban/rural classifications), misreporting (due partly to the fears caused by widespread xenophobic sentiments among the population and an inflexible immigration policy), and social as well as spatial issues (that can be ascribed to the legacy of apartheid).

Purpose-made migration surveys, although better suited to overcome some of the limitations of census-based migration data, also have serious limitations. Sample surveys lack the spatial detail to undertake analyses of migration patterns in any depth, and they may not be sufficiently representative of the population to enable researchers to draw firm conclusions about international (and especially cross-border) migration issues. Qualitative research, although very important for obtaining insights into the underlying phenomena, processes, causes and consequences, usually provides no basis for generalisation.

⁴ The internationally comparative Migration and Health Survey conducted by Pretoria University in 2000 is another potentially important data source, and can be accessed at <http://www.nrf.ac.za/sada/index.asp>.

⁵ The local-level, longitudinal data generated by the Agincourt DSS (Demographic Surveillance Site) should be regarded, amongst others, as another potentially important information source.

⁶ Unfortunately the only migration data from Census 2001 that have so far been made publicly available – as far as I know – are those contained in the 10% sample and the first release of the Migration Community Profile, which unfortunately covered only inter-provincial migration. However, a comprehensive Migration Community Profile data set will hopefully be made available soon.

DATA AVAILABLE

The population censuses of 1996 and 2001 provided us with very important and useful migration, spatial and other data for deriving the necessary context for migration analyses. The opportunity presented by the data from these two censuses should now be grabbed with both hands. Data are also available from national migration and other surveys (e.g. the HSRC's 2000 and 2001–02 migration surveys). These data are invaluable for more specialised migration studies.

Migration studies can also benefit from longitudinal surveys for more in-depth research. The longitudinal research undertaken at the various Demographic Surveillance Sites, such as Agincourt DSS and by the Africa Centre are good examples. The proposed national longitudinal study by the HSRC is another potentially important source of data and migration information.

QUALITATIVE RESEARCH

Qualitative research is necessary for understanding the underlying processes causing migration and non-migration. It is also important to link qualitative work with survey research to understand the 'true meaning' of survey responses. Unfortunately this was not done with a view to complementing the two recently completed HSRC migration surveys. I have come to realise that this was an important oversight on my part.

CONCEPTUAL FRAMEWORK AND HYPOTHESES

Two national surveys were undertaken in the HSRC's migration project. The first (initial or preliminary) survey, among 911 respondents, was conducted with a view mainly to test certain items for the purposes of the second (main) survey, which covered 3 618 respondents.

Based on the theory of planned behaviour of Fishbein and Ajzen (1975) and Ajzen (1988) and its application to migration theory by De Jong and Fawcett (1981) and De Jong (2000), it can be hypothesised firstly that *place-related expectations – the process of evaluating the chances for future attainment of valued goals in the home community (stay decision) versus an alternative location (move decision) – along with migrant networks, family influences and information flows, are a primary determinant of intentions to move or stay, which in turn are a proximate determinant of migration in South Africa today.*

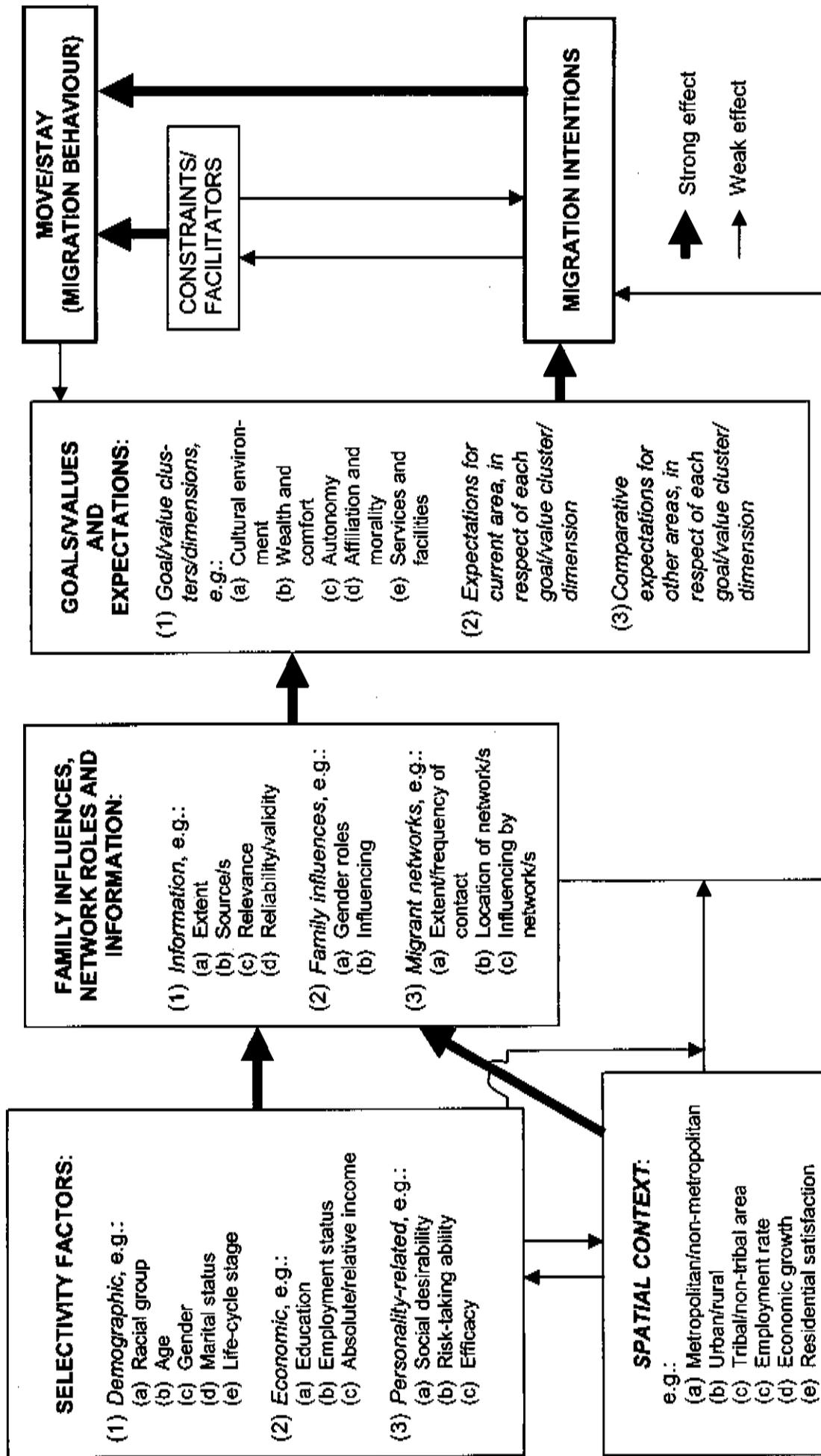
The key variables needed for testing this hypothesis will be actual migration behaviour (to be obtained from the proposed longitudinal survey), and the following variables obtained from the 2001–02 HSRC Migration Survey: (a) migration intentions expressed during the survey, (b) the six goal (value) dimensions that were confirmed by the survey, (c) the comparative expectations for valued goals to be realised in the envisaged possible destination (relative to the expectations in respect of the area where the main survey took place), versus (d) expectations for the same goals to be attained in the area where the main survey took place.

The value-expectancy formula to be used in these analyses, which is a modification of the formula suggested by De Jong and Fawcett (1981:47), will be as follows:

$$MI = \sum_{i=1}^k V_i * E_{2i} / E_{1i}$$

- where:
- MI = the strength of the intention to migrate
 - V_i = value attached to a particular goal item (i)
 - E_{1i} = expectation for the current area of residence, denoting the extent to which the particular goal item (i) is likely to be achieved in the current area of residence
 - E_{2i} = comparative expectation in respect of the possible destination, denoting the extent to which the particular goal item (i) is likely to be better or worse achieved in the possible area of destination than in the current area of residence
 - k = number of goal items included in the analysis
 - $*$ = multiplication sign

FIGURE 1: SUGGESTED FRAMEWORK FOR AN EXPECTANCY-BASED MODEL OF MIGRATION DECISION MAKING



The value (V) component in the above formula can be regarded as a mechanism for "weighting" the expectancy (E) components in the formula in terms of the differential importance attached to each of the goal items (i). Expectations are thus the key components for the purposes of the proposed analyses.

In the analyses for the hypothesis testing extensive use will be made of structural equation (path) models of longitudinal event-history migration data. The procedure for testing this model will be to design a set of structural equations that reflect the relationships depicted in Figure 1.⁷ The aims of these statistical analyses will be to provide unique answers concerning the causal order of the variables included in the conceptual framework.

It can further be hypothesised that *migration intentions and behaviour, as well as expectations, migrant networks, family influences and information flows, are differentiated by spatial context and racial group, because of the historical and current area-level and individual-level racial stratification systems and development policy in South Africa.* The data needed for testing this hypothesis will be obtained mainly from the proposed longitudinal study and the 2001–02 HSRC Migration Survey, but also utilising migration and other data from Census 1996 and Census 2001. As indicated in Figure 1 the analyses will also take into account the demographic and economic characteristics of the respondents and their personal traits, the spatial context (which not only covers micro-level data such as satisfaction with life on the whole, but also macro-level data on regional economic growth and meso-level economic data on employment rates and income levels that will be obtained from Statistics South Africa).

The aims of these analyses will be to provide unique answers concerning the effects of multi-level factors on migration intentions and behaviour. The emphasis in the testing of the second hypothesis will therefore be to take the analyses for the testing of the first hypothesis one step further by dealing specifically with the need to properly analyse the multi-level effects.

LONGITUDINAL APPROACH

In the 2001–02 HSRC Migration Survey detailed contact particulars of the respondent were obtained (if s/he agreed to being interviewed during a follow-up survey). An attempt was also made to collect contact information for the respondents' spouses, their adult children, other close relatives and also close friends. It was, however, necessary to first determine the degree of potential bias, resulting from refusals to be re-interviewed or respondents not being traced. In an attempt to determine the viability of the proposed longitudinal survey, a feasibility survey was undertaken during June 2004. The aim of the feasibility survey was to calculate the probability of tracing the original respondents and confirming their contact particulars or get their new details.⁸ During the feasibility survey it was attempted to contact all the households covered in the 10 selected EAs.

The feasibility survey showed that the proposed longitudinal survey would be viable. Of the 59 respondents of the main (2001–02) survey 48 (81%) were traced, four (7%) had died since the main survey, and only seven (12%) could not be located at all.⁹ With only five (9%) of the respon-

⁷ Graphical chain models could be applied, for example, to supplement the path analyses. These models allow for log-linear and logistic models to be applied in cases of categorical responses, resulting in the use of the proper model to study each of the pathways (see, for example, Mohamed, Diamond & Smith 1998; Smith 2003, and Magadi et al. 2004).

⁸ For the purposes of this survey a sample of 10 EAs was drawn, based *firstly* on the original stratification (but combining some strata), *secondly* limiting the survey to the four northern provinces plus KwaZulu-Natal (so as to save costs), and *thirdly* ensuring coverage of all five fieldwork companies used in the 2001–02 survey. The geographic distribution of the sample for the feasibility survey was as follows: (a) Gauteng: 2 EAs (both urban; in two magisterial districts), (b) KwaZulu-Natal: 3 EAs (all urban; in two districts), (c) Limpopo: 1 EA (rural), (d) Mpumalanga: 2 EAs (one urban, one rural; in two districts), and (e) North West 2 EAs (one urban, one rural; in two districts).

⁹ Of the 55 respondents from the main survey who were still alive at the time of the feasibility survey, 39 (70%) were still living at the very same address, while seven (13%) had moved to other addresses but still in the same area (therefore constituting residential mobility; i.e. not migration as such). Three persons had moved away to unknown addresses (presumably in other areas), with one having moved abroad. A further

dents having moved away to other areas, there has clearly been a low level of migration among the respondents, thereby confirming the conclusions of Anderson (2003) and Kok et al. (2003) that the post-apartheid migration levels were for all practical purposes as low as they had been during the height of apartheid.

METHODOLOGICAL ISSUES

Two further methodological issues facing the proposed longitudinal study, namely rationalisations and the appropriate construction and treatment of variables, are briefly introduced here.

(a) Rationalisations

The problem of rationalisation is an important issue when trying to plumb the motives underlying migration intentions and behaviour. Rationalisation should be a particularly salient concern in research on motives for past behaviour, where post facto rationalisations tend to dominate responses (Bedford, quoted in De Jong & Fawcett 1981:44), but it also applies to possible future migration (albeit perhaps in a somewhat different form). Gelderblom (2003a) correctly points out that because the failure to pursue better opportunities elsewhere has the connotation of non-migrants being regarded as 'losers', they might be inclined to rationalise their inability to migrate. It may be true that there is no simple way of overcoming this problem (Bedford, in De Jong & Fawcett 1981:44), but it is suggested that by controlling for different people's differential social desirability needs (e.g. in multivariate analyses), one can probably reduce this effect.

(b) Variable construction and treatment

The appropriate construction of the variables relating to expectations, migrant networks, family influences, information, spatial context and selectivity factors requires specific attention. It will not be necessary to deal here with the variables concerned, but it is important to mention one of the pitfalls: too often proxies are used to infer variables, and sometimes only some elements of a concept are used to construct variables. This means that complex concepts should first be theoretically evaluated and unpacked before a construction of appropriate variables is attempted.

In the proposed longitudinal research particular attention is given to the construction of appropriate variables. This applies especially to the key variables that will be investigated, namely migrant networks, family influences, information flows and selectivity factors as they relate to expectations, migration intentions and subsequent behaviour.

(i) Expectations

Gelderblom (2003a), in his critique of the value-expectancy framework, gives credit to the fact that "some space is given for individuals to state their own preferences", but he is critical of the fact that "...this is limited to a rating of different places relative to a number of preset values...". While also giving recognition to the fact that the value dimensions were identified by De Jong and Fawcett (1981) on the basis of "an analysis of the reasons given by respondents for their migration behavior as these emerged in surveys, as well as migration theories", Gelderblom (2003a) is concerned about the fact that these were then reduced to the seven "conceptual categories that seem to represent psychologically meaningful clusters" (De Jong & Fawcett 1981:49). It is therefore necessary to also consider other, perhaps non-psychological, dimensions (clusters) when trying to capture the widest possible range of value-expectancy dimensions.

Based on the exploratory factor analysis and the subsequent item analyses on the data from the initial (preliminary) HSRC migration survey (in 2000), a total of 10 value dimensions, ranging from psychological to environmental and service-delivery factors, were

four (7%) of the originally recorded physical addresses could not be found again, therefore making it impossible to trace the original respondents.

expected to be associated with migration in the main survey (of 2001–2002). As it turned out, four of these had to be dropped because they could not pass the requirements of the confirmatory factor analysis. The remaining six clusters (“cultural environment”, “wealth and comfort”, “autonomy”, “affiliation and morality”, and “services”) were found suitable for the purposes of the study, and can be used to study value-weighted expectations.

(ii) Migrant networks

In De Jong's (2000) report on the study in Thailand it is stated that migrant networks in Bangkok and regional urban locations are not significant predictors of migration (be it temporary or more permanent) behaviour. Respondents had been asked to assign a value of “1” for the presence of one or more (a) relatives and (b) friends in Bangkok, the provincial center and the regional center respectively. These responses were then summed to obtain a score of between zero and six. Gelderblom (2003b) has the following to say about this approach: “Even though one can make a logical distinction between the dual role of networks in making migration more likely and directing it to particular destinations, for the purposes of empirical study there is no clear cut distinction. A migrant network affects the likelihood of migration in general only through its impact in increasing the likelihood of moving to a particular place, and that place is the place where the aspiring migrant has ‘auspices’. ... It is therefore wrong to measure network connections in different places in an additive way.”¹⁰

(iii) Family influences

The items used in De Jong's (2000) treatment of “family migration norms” can perhaps be extended in an attempt to increase the reliability of the variable. De Jong measures family migration norms by the respondent's perception of whether his/her mother, father, spouse or child has encouraged him/her to stay or move in the preceding two years, and the categories he uses are (1) “encourage to stay”, (2) “not stated”, and (3) “encourage to move” – for the first available response for mother, spouse, father or child. It may be advisable to include other family members (apart from mother, father, spouse or child), especially in contexts such as southern Africa where extended families and complex household structures are the norm rather than the exception. By taking into account the number of persons regarded as very important in the life of the respondent (i.e. “significant others”) who had recently attempted to influence the respondent to either move or stay respectively, a somewhat more nuanced treatment of the effects of “family influences” may be possible.

(iv) Selectivity factors

The appropriate treatment of selectivity factors in migration models is another methodological issue to be considered. Lucas (1997:728) emphasises the fact that “ultimately one wishes to know why migration is selective upon particular characteristics”. He

¹⁰ In the 2001–02 HSRC Migration Survey the following two questions were asked about contacts in other possible destinations: (a) “Do you have any immediate relatives or close friends who live [there]?”, and if the answer was “yes”, the respondent was asked (b) “Have you or other members of this household had contact with any of these relatives or friends living [there] during the past 12 months?” A simple network variable can then be constructed by giving it the value ‘1’ if the response to both questions is affirmative or the value zero if any one of the questions is answered in the negative.

A third question asked in the HSRC survey relates (qualitatively) to the nature of the contact. If the response to the second question above is “yes”, the respondent is asked: “What was the nature of the contact you or other members of this household had with the relatives or friends living in Johannesburg during the past 12 months?” The multiple response options to this question were: (a) “We/I visited them [there]”, (b) “They visited us/me”, (c) “We/I talked to them on the phone”, (d) “We exchanged personal letters”, (e) “They offered one of us a job [there]”, (f) “They provided information about jobs/work [there]”, (g) “They offered us/me money or gifts”, and (h) “They gave us/me money or gifts”. By constructing an ordinal-scale variable denoting the increasing potential impacts of the networks, this will allow a more nuanced reflection of the functioning of networks.

points to the need for incorporating selectivity factors in structural equation models, because these factors cannot be determined from a single-equation, reduced-form analysis of regression upon personal characteristics alone: "A structural, behavioral model is required. Indeed within such a framework, many personal characteristics – such as the extent of schooling and marital status – are endogenously determined together with the migration decision" (p. 728).

There is no need to discuss the construction of variables to deal with demographic factors such as age and sex, but the selectivity variable "life-cycle stage" perhaps warrants some unpacking. Haley (1973), who is said to have introduced the term "family life cycle" into the psychiatric literature, identified as many as eight stages in the lives of married couples, each with its own important demands and with every transition from one stage to the next being potentially migration-inducing. However, seldom in the migration literature would these stages be reported or at least be accounted for, which serves to illustrate the need to cover all relevant aspects of the selectivity and other concepts used in the study.

Furthermore, as De Jong and Fawcett (1981) correctly point out, research on risk taking and similar personal traits is often flawed by the use of education or some other proxy measures for personal traits. "Because of the paucity of studies that measure traits directly, the strength of personal traits as determinants of migration decision making is unclear. Our working hypothesis is that individuals whose perception of themselves include personal efficacy, adaptability to change, and the ability to take risks are more likely than others to express values and expectancies favoring spatial mobility" (De Jong & Fawcett 1981:55).

A superficial treatment of the set of selectivity variables called "personal traits" must therefore be avoided. Here we'll need to focus our attention on three of these variables, namely social desirability, the ability to cope with risks, and efficacy, with the latter two having been identified as potentially important determinants of migration (albeit perhaps indirectly via expectancy, as suggested by De Jong and Fawcett 1981). Yet not one of the publications quoted earlier indicates what exactly these variables are or how they should be constructed.

The issues of rationalisations and socially desirable responses were mentioned earlier are perhaps a case in point. At least a partial solution may be found in the HSRC migration surveys – on which the proposed longitudinal study will be based – where the six items of the shortened social desirability scale by Greenwald and Satow (1970)¹¹ were included. These items have since been subjected with some success to confirmatory factor analysis and item analysis (see Appendix 1, where a brief description of the findings is given). It should therefore be possible to control for the potential effects of socially desirable responses in the research proposed here.

In Appendix 1 the findings of the factor and item analyses (from the main 2001–02 survey) and the results of reliability and validity tests (from the initial 2000 survey) are given mainly in respect of the above-mentioned two selectivity factors, namely "risk-taking ability" and "efficacy".

The analytical opportunities offered by the data from Census 1996 have perhaps already been fully explored in a recent publication by the HSRC (see Kok et al. 2003). Statistics South Africa is currently planning a monograph that should do much the same for the migration data generated by Census 2001, and the HSRC also plans to publish, in collaboration with Statistics SA, a web-based users handbook for utilising the latest census data.

¹¹ This shortened scale was based on the more comprehensive Crowne-Marlowe (1964) social desirability scale, and has been applied successfully in South Africa by, among others, Duckitt and Broll (1982) and Kok (1988).

These initiatives may not be enough, though. There is a case to be made for showing how sample-survey data and information generated by qualitative research can be used to augment and help interpret census data. The two HSRC surveys, albeit theoretical in aim and attempting specifically to determine the micro-level dynamics and causes of migration, probably go a long way toward achieving this, but these need to be used in conjunction with longitudinal and qualitative research as well as meso-level and macro-level data obtained from other sources, including censuses.

CONCLUSIONS

Migration analysis is complex. Nevertheless, the need for a proper understanding of migration processes in this part of the world requires that these analyses be undertaken. Researchers that are trained and experienced in demographic analysis should become involved in the required studies. Our universities should produce such analysts, and government should provide them with the opportunities to gain essential experience in research of that nature.

The absence of useful migration data was one important reason why South Africa historically did not produce sufficient numbers of demography-trained migration analysts. This situation needs to be rectified sooner rather than later. Fully utilising the opportunities offered by our recent censuses and migration surveys, while effectively dealing with the problems inherent to migration data, is therefore of crucial and immediate importance.

Data generated in longitudinal migration research are particularly appropriate for this purpose, and the paper suggests that two recent HSRC migration surveys have, for the very first time in South African history, laid a foundation suitable for such longitudinal analytical research. However, a serious attempt should be made during the proposed HSRC longitudinal study to at least partially rectify the past mistake of not complementing the surveys with linked qualitative research.

On the one hand, it is now up to universities and researchers to heed the call by government for appropriate migration information to inform policy making and planning. Government, on the other hand, has the responsibility to fund the research needed to provide the required information.

The methodology proposed here clearly has its potential limitations. Firstly, the migration data from Census 2001 have as yet not been fully evaluated. To suggest, therefore, that the latest census provides an opportunity for analysing migration trends over time may thus be somewhat premature. Secondly, the serious shortcoming of the HSRC's sample-based migration surveys was the absence of appropriately linked qualitative research that would have provided very important complementary insights into the dynamics and causes of migration. This oversight must be rectified during the proposed longitudinal study.

The theme of this Conference is "Towards Sustainable Population and Development". The paper attempts to provide a framework for achieving this goal in the field of migration research. It provides a potentially useful theoretical foundation, based on our current understanding of migration, and also offers methodology-based suggestions for building further on that foundation.

REFERENCES

- Ajzen, I. 1988. *Attitudes, personality and behavior*. Milton Keynes: Open University Press.
- Anderson, B A. 2003. *Migration in South Africa in comparative perspective*. Paper read at the HSRC Migration Workshop, Pretoria, 17–20 March 2003.
- Crowne, D P & D Marlowe. 1964. *The approval motive*. New York: Wiley.
- De Jong, G F. 2000. "Expectations, gender, and norms in migration decision-making". *Population Studies*, 54(3):307–319.
- De Jong, G F & J T Fawcett. 1981. "Motivations for migration: An assessment and a value-expectancy research model". In: De Jong, G.F. & Gardner, R.W. (eds). *Migration decision making: Multidisciplinary approaches to microlevel studies in developed and developing countries*. New York: Pergamon, pp. 13-58.
- Duckitt, J H & T Broll. 1982. "Life event stress, social support and health". *Humanitas*, 8(4):377–383.
- Fishbein, M & I Ajzen. 1975. *Belief, attitude, intention and behavior: An introduction to theory and research*. Reading, Mass.: Addison-Wesley.
- Gelderblom, D. 2003a. *A theoretical model of migration*. Paper read at the HSRC Migration Workshop, Pretoria, 17–20 March 2003.
- Gelderblom, D. 2003b. *Social networks in the migration process*. Paper read at the HSRC Migration Workshop, Pretoria, 17–20 March 2003.
- Greenwald, H J & Y Satow. 1970. "A short social desirability scale". *Psychological Reports*, 27: 131–135.
- Haley, J. 1973. *Uncommon therapy: the psychiatric techniques of Milton H. Erikson*. New York: Norton.
- Jackson, D N. 1976. *Jackson personality inventory manual*. Goshen (NY): Research Psychologists Press.
- Kok, P C. 1988. *Migrasiebesluitneming: 'n Empiriese toepassing van die waardeverwagtingsmodel op Suid-Afrikaanse blankes* [English translation: *Migration decision making: An empirical application of the value-expectancy model on South African whites*]. Verslag S-180. Pretoria: RGN.
- Kok, P, M O'Donovan, O Bouare & J van Zyl. 2003. *Post-apartheid patterns of internal migration in South Africa*. Cape Town: HSRC.
- Kok, P & J Pietersen. 2003. *Migration causes in South Africa: Confirmatory factor and item analyses on data from a sample survey*. Unpublished report. Pretoria: HSRC.
- Lucas, R E B. 1997. "Internal migration in developing countries". In: M R Rosenzweig & O Stark (eds). *Handbook of population and family economics (Volume 1B)*. Amsterdam: Elsevier, pp. 721–798.
- Magadi, M, I Diamond, N Madise & P W F Smith. 2004. "Pathways of the determinants of unfavourable birth outcomes in Kenya". *Journal of Biosocial Science*, 36, 153–176.
- Mohamed, W N, I Diamond & P W F Smith. 1998. "The determinants of infant mortality in Malaysia: A graphical chain modelling approach". *Journal of the Royal Statistical Society. Series A*, Vol. 161, No. 3. (1998), pp. 349-366.
- Schwarzer, R. 1998. *General perceived self-efficacy in 14 cultures*. (Internet URL: <http://www.userpage.fu-berlin.de/~health/world14.htm>. Last accessed on 24 October 2003.)

- Schwarzer, R & M Jerusalem. 1993. *The general perceived self-efficacy scale*. (Internet URL: <http://www.userpage.fu-berlin.de/~health/engscal.htm>. Last accessed on 24 October 2003.)
- Smith, P W F. 2003. "Graphical modeling". In: M Lewis-Beck, A E Bryman & T F Liao (eds). *The SAGE Encyclopedia of Social Science Research Methods*. Sage Publications.
- Trochim, W.M.K. 1999. *Research methods knowledge base*. 2nd Edition. (Internet URL: <http://trochim.human.cornell.edu/kb/constval.htm>. Updated [2002] version accessed on 24 October 2003.)

APPENDIX 1

RELIABILITY AND VALIDITY OF ITEMS RELATED TO "PERSONAL TRAITS"

Since the two personality dimensions "risk-taking ability" and "efficacy" have been identified in the literature as important for migration research, they warrant a full investigation. A number of possible items that had been hypothesised to contribute significantly to the measurement of these two dimensions (partly borrowed from the "risk-taking scale" by Jackson 1976, and the "self-efficacy scale" by Schwarzer & Jerusalem 1993 respectively¹²) were included in the questionnaire for the HSRC's initial (2000) survey and a reduced number, based on the analyses on the data from the initial survey, was used in the main (2001–02) HSRC Migration Survey.¹³ Factor and item analyses were conducted on these items following each survey, and the findings are presented below.

The question that dealt with these attitude items read as follows in both surveys: "I shall now read you a number of statements. These statements reflect certain people's attitudes to specific matters. Please tell me to what extent you agree or disagree with each particular statement." The items used in this battery of questions were measured on a Likert-type, five-point scale with the following response options: (a) "strongly disagree" (with a value of 1), (b) "disagree" (value = 2), (c) "neither agree nor disagree / uncertain" (3), (d) "agree" (4), and (e) "strongly agree" (5).

There was also an option for the respondent to indicate that s/he did not understand the particular item. Such responses were treated as missing values in the analyses. The "don't understand" option was included to identify those items that may be less suitable for the purposes of the study if a significant proportion of respondents would not understand the meaning of the item. No single item included below had a notable proportion (1% or more) of "don't understand" responses.

INITIAL (2000) SURVEY

Risk-taking propensity

A battery of 20 potential items to measure risk-taking propensity was included in the questionnaire of the initial survey. Of these only the eight below were found sufficiently acceptable (including not being culture-specific, which proved to be a major problem). This item analysis was based on the responses of 707 individuals.

- (1) "If the possible reward was very high, I would not hesitate putting my money into a new business, even though it could fail" [Factor loading (FL)¹⁴ = 0.79; Discrimination value (DV)¹⁵ = 0.51]
- (2) "If the gains involved are high, it does not bother me to take risks" [FL = 0.66; DV = 0.59]
- (3) "People have told me that I enjoy taking chances" [FL = 0.65; DV = 0.53]
- (4) "I enjoy the challenge of a project that could mean either a good promotion or the loss of my job" [FL = 0.75; DV = 0.43]

¹² It should be understood that the diversity of the South African population makes it very difficult to find items appropriate for all our cultural and socio-economic contexts. This highlights the need for a detailed testing of questionnaire items of this nature.

¹³ It should be made clear from the outset that the two suggested scales were not meant to be used here as psychometric tests in the usual sense. *Neither the risk-taking nor the efficacy scale is intended for individual psychological assessment.*

¹⁴ The factor loading denotes the results of the rotated (based on an oblique rotation using the "promax" rotation method of SAS) factor pattern, i.e. the degree of association between each observed item and the specific factor. The standardised regression coefficients obtained from the factor analysis are reported here.

¹⁵ The discrimination value (DV) indicates the extent to which the particular item distinguishes between individuals who attained a high score on the scale and those who attained a low score. The DVs presented here were adjusted for the number of items in the particular scale.

- (5) "I am the kind of person who likes to take risks" [FL = 0.66; DV = 0.44]
 (6) "I never take risks when there is another alternative" [FL = -0.47; DV = 0.27]
 (7) "When I want something, I'll sometimes 'go all out' to get it" [FL = 0.47; DV = 0.26]
 (8) "I prefer spending my time in my own, familiar environment" [FL = -0.13; DV = 0.12]

The reliability coefficients¹⁶ (obtained from the item analysis) for these eight items are as follows:

- (a) KR20: 0.70¹⁷
 (b) KR8: 0.79¹⁸

Reliability is not the only criterion for a good scale, however. Although reliability (i.e. the ability of a measuring tool to produce the expected outcome repeatedly) is a necessary condition, it is not sufficient for assuming *validity* (i.e. the requirement that a tool must measure what it is supposed to measure). Consequently, it is possible for a test to be reliable without being valid (i.e., a test can give the same result time after time but not be measuring what it was intended to measure). Validity should therefore be determined as well.

A series of validity tests is needed to determine whether a particular tool (scale) is valid or not. There are essentially three types of validity: (a) internal validity (covering both "face validity" and "content validity"), (b) external validity (which covers both forms of "criterion-related validity"), and (c) construct validity. These will be discussed separately as they are applied to the suggested scale on risk-taking propensity.

* This item was reverse-coded as follows: $N = 6 - O$ (where N = new value, and O = original value).

¹⁶ "Reliability" refers to the extent to which scale scores are consistent; i.e. the degree to which the scale scores are dependable or relatively free from random errors of measurement. Reliability therefore indicates the extent to which a measure consistently produces the same result under similar conditions. An important element of reliability is inter-item consistency, which is the extent to which all the items on the test are measuring the same thing. Some of the factors contributing to a *high* level of reliability are the following: (a) the number of items; (b) the homogeneity of the items; (c) the heterogeneity of the sample, and (d) high discrimination values (DVs) of the items.

Reliability is usually expressed in the form of a reliability coefficient. The *reliability coefficients* reported here were estimated by calculating the correlation through procedures known as internal-consistency estimates. These measure the degree of homogeneity of content on multi-item scales. For a scale to be internally consistent there must be high correlations among the items. Some of the most commonly used estimates of reliability in item analyses are the Kuder-Richardson formulae (described below).

¹⁷ The KR20 (short for Kuder-Richardson Formula 20) is an overall measure of internal consistency, sometimes also referred to as an overall scale index of item homogeneity. The KR20 has a normal range between 0,00 and 1,00 with higher numbers indicating higher internal consistency. The KR20 is sensitive to the number of items, with a scale involving more items receiving an elevated score. The KR20 is calculated by means of the following formula:

$$KR20 = \left[k / (k - 1) \right] \left\{ (S^2 - \sum_{i=1}^k S_i^2) / S^2 \right\}$$

where: k = number of items
 S^2 = observed variance of the entire factor (i.e. over all items)
 S_i^2 = observed variance of the particular item (i)

¹⁸ The KR8 (Kuder-Richardson Formula 8) coefficient is also a measure of internal consistency. Unlike the KR20 coefficient, KR8 takes into account the discrimination values of the individual items, and does not necessarily increase as the number of items increases. The KR8 coefficient is calculated as follows:

$$KR8 = \left\{ \left(S^2 - \sum_{i=1}^k S_i^2 \right) / 2 * S^2 \right\} + \sqrt{\sum_{i=1}^k DV_i^2 S_i^2} + \left\{ (S^2 - \sum_{i=1}^k S_i^2) / 2 * S^2 \right\}$$

where: k = number of items
 S^2 = observed variance of the entire factor (i.e. over all items)
 S_i^2 = observed variance of the particular item (i)
 DV_i^2 = square of the discrimination value of Item i

(a) *Internal validity*

Internal validity indicates the degree to which the researcher is able to understand the data and draw conclusions from them.

(i) *Face validity*

The most commonly known version of content validity is "face validity", which basically refers to the degree to which a test appears to be measuring what it purports to measure. Although not a sound way of estimating validity, it is often used only to determine whether the items selected can be expected to form a valid construct. When planning the battery of items to be included, the researcher would normally expect to find out what s/he wants to know by asking that particular set of questions. The face validity of the above scale for risk-taking propensity should be clear from the items listed above.

(ii) *Content validity*

Content validity is a measure of the degree to which a scale measures an intended content area. As such it requires both "item validity" and "sampling validity". *Item validity* indicates whether the scale items represent measurement in the intended content area, while *sampling validity* indicates how well the scale samples the total content area. No formula for computing it is available, and it cannot be expressed in quantitative terms. Content validity can therefore be determined only by informed judgment, based on the literature or on expert opinion.

A majority of the items listed above were taken from (or based on) the Jackson personality inventory (1976) on risk-taking ability. Furthermore, some of the items included in the suggested scale had been included in the scale on risk-taking propensity applied by Kok (1988) in his study of internal migration among South African whites. Consequently, it would be correct to say that the suggested scale seems to have sufficient content validity to warrant its application in the envisaged further migration research.

(b) *External validity*

External validity generally indicates the degree to which the researcher is able to generalise the findings of his/her research. External validity is therefore the degree to which the conclusions in a particular study would hold for other persons in other places and at other times. The two forms of criterion-related (external) validity are "concurrent" and "predictive" validity.

(i) *Concurrent validity*

Concurrent validity refers to the degree to which the scores on a scale are related to the scores on another, already established scale administered at the same time or to some other valid criterion available at the same time. Often a scale or test is developed which claims to do the same job as some other scale, but more easily or faster.¹⁹ Since no attempt was made to administer any other scale at the same time, the concurrent validity of the suggested scale for risk-taking propensity could not be determined.

(ii) *Predictive validity*

Predictive validity indicates the degree to which a scale can predict what an individual is likely to do in a future situation. In predictive validity, one assesses the scale's *ability to predict something it should theoretically be able to predict*.

In this case it is hypothesised that risk-taking propensity would be a good predictor of the probability of migration to have occurred in the past or to occur in future. It would therefore be important to determine the extent to which the suggested scale of risk-taking propensity succeeds in explaining the following:

¹⁹ The shortened social desirability scale by Greenwald and Satow (1970) is an example of such a scale.

- whether or not the respondent has ever migrated voluntarily since reaching the age of 18 years
- the respondent's duration of stay at the current place of residence
- whether or not he/she intends to migrate in future

A simple Pearson correlation analysis indicated, however, that risk-taking ability was *not* significantly correlated (at the 5% level) with past migration, duration of stay or migration intentions. A series of regression analyses did provide significant results, though. These are presented below.

(aa) Risk-taking propensity and previous voluntary migration

A logistic regression produced the following equation (including social desirability, age and education as control variables):

$$\text{logit}(p) = -2.96 + 0.015x_1 + 0.004x_2 + 0.059x_3 - 0.338x_4$$

where: p = probability of having ever migrated voluntarily during adulthood
 x_1 = risk-taking propensity (expressed as a percentage)²⁰
 x_2 = social desirability (expressed as a percentage)²¹
 x_3 = age (in single years)²²
 x_4 = index of educational attainment²³

This model correctly classified 61% of the 658 observed cases used in the analysis, and all the coefficients are significant at the 0.01% level.

(bb) Risk-taking propensity and duration of stay at the current place of residence

An ordinary least-squares regression (using the same control variables as before) produced the following equation:

$$y = 7.01 - 0.057x_1 - 0.121x_2 + 0.577x_3 + 1.725x_4$$

where: y = duration of stay at current place of residence (years)
 x_1 = risk-taking propensity (expressed as a percentage)²⁴
 x_2 = social desirability (expressed as a percentage)
 x_3 = age (in single years)
 x_4 = index of educational attainment

This model explains 23% of the variation in duration of stay, and is significant at the 0.01% level. All coefficients in the model are significant (at least at the 8% level).

(cc) Risk-taking propensity and migration intentions

A logistic regression (using the same control variables as before) produced the following equation:

$$\text{logit}(p) = -2.28 - 0.008x_1 + 0.009x_2 - 0.027x_3 + 0.422x_4$$

where: p = probability of considering to migrate in future

²⁰ The percentages for the scales used here were calculated as follows:

$$\%score = (\text{raw score} - \text{minimum possible score}) / (\text{maximum possible score} - \text{minimum possible score}) \times 100$$

²¹ The correlation analysis showed that risk-taking propensity was correlated significantly with social desirability ($r = -0.24$). This indicates that one should at all times control – in a multivariate context – for the effect of social desirability when utilising this scale of risk-taking propensity.

²² The ages of the respondents ranged from 18 to 69 years.

²³ This index (measured at an ordinal scale ranging from 0 to 5) roughly represents the number of years of education (including both school and post-school education).

²⁴ The negative sign of the coefficient for risk-taking propensity (when modelling duration of stay) is as expected. The lower the respondent's score on the risk-taking propensity scale, the longer s/he would be expected to have "stayed put" in the current place of residence.

X_1	=	risk-taking propensity (expressed as a percentage)
X_2	=	social desirability (expressed as a percentage)
X_3	=	age (in single years)
X_4	=	index of educational attainment

This model correctly classified 64% of the 625 observed cases used in the analysis, and all the coefficients are significant at the 0.01% level.

In all cases but one the signs of the regression coefficients were as expected. The one exception is the sign of the regression coefficient for risk-taking propensity in the last equation. This coefficient is negative, indicating that respondents with a higher risk-taking propensity were less inclined to consider migrating in future. Although this may seem to be a somewhat disappointing finding, it is not necessarily a problem. In terms of De Jong and Fawcett's (1981:54) causal framework, risk-taking propensity (as one of the background variables in the category "personal traits") operates predominantly through intermediate variables and therefore mainly affects migration intentions indirectly. More substantial multi-variate analyses should be used to corroborate that theoretical proposition, though.

The predictive validity of the suggested risk-taking propensity scale has therefore largely been established.

(c) Construct validity

Construct validity refers to the degree to which a test or scale measures an intended hypothetical construct.²⁵ As such, construct validity deals with relationships, and involves an understanding of the theoretical rationale underlying the phenomena to be measured. Like external validity, construct validity is related to generalisation, but where external validity involves a generalisation from the particular study context to other people, places or times, construct validity involves a generalisation from a particular measurement to the *concept* of the measurement. One way of conceptualising construct validity is therefore to view it as a "labeling" issue (Trochim 1999).

Trochim (1999) goes as far as suggesting that construct validity should be viewed as the overarching quality with all of the other measurement validity labels falling beneath it. According to him, construct validity is as much a part of the independent variable as it represents the dependent variable. When researchers claim that their tools (scales in this case) have construct validity, they are essentially claiming that they understand how their constructs operate in theory, and that they can provide evidence that these constructs behave in practice the way they think they should. For construct validity one is essentially claiming that the observed pattern – how things operate in reality – corresponds with the theoretical pattern. Trochim (1999) calls this process "pattern matching", and believes that this is at the heart of construct validity. He states further that, in order to establish construct validity, one has to meet the following conditions:

- (i) One has to set the construct one wants to operationalise (e.g. "self-esteem") within a *semantic net* (or "net of meaning"). This means that one has to indicate what the construct is more or less similar to in meaning.
- (ii) One needs to be able to provide direct evidence that one *controls* the operationalisation of the construct – that one's operationalisation confirms the theoretically expected patterns. (If one is trying to measure self-esteem, for example, one has to be able to explain why one operationalised the questions the way one did. For example, if all of one's questions are addition problems, how can it be argued that one's measure reflects "self-esteem" and not "adding ability"?)
- (iii) One has to provide evidence that one's *data support one's theoretical view* of the relations among constructs. For example, if it is believed that self-esteem is closer in meaning to "self-worth" than it is to "anxiety", one should be able to show that measures of self-esteem are more highly correlated with measures of self-worth than with ones of anxiety.

²⁵ A "construct" is a non-observable trait that explains behaviour. One cannot observe a construct (such as "risk-taking propensity") – only its *effect* can be observed. Constructs are thus "invented" in an attempt to explain behaviour.

All three of Trochim's (1999) conditions have therefore been met to a large extent (as discussed under "internal validity" and "external validity" above).

The reliability and validity of the risk-taking ability scale presented here have therefore been confirmed for the most part.

Efficacy

General perceived self-efficacy pertains to optimistic beliefs about being able to cope with a large variety of stressors. In contrast to other constructs of optimism, perceived self-efficacy explicitly refers to one's competence to deal with challenging encounters (Schwarzer 1998:1).

The "Generalised Self-Efficacy Scale" was originally developed in German by Matthias Jerusalem and Ralf Schwarzer in 1981 and has since been used in many studies with a large number of participants (see <http://userpage.fu-berlin.de/~health/world14.htm>). It consists of a 10-item psychometric scale that was designed to assess optimistic self-beliefs to cope with a variety of difficult demands in life. However, "in contrast to other scales that were designed to assess optimism, this one explicitly refers to personal agency, i.e., the belief that one's actions are responsible for successful outcomes" (Schwarzer 1998).

However, of the 10 potential efficacy items originally included in the questionnaire, only the following eight survived the factor and item analyses based on the initial survey:

- (1) "When I am confronted with a problem, I can usually find several solutions" [Factor loading (FL) = 0.71; Discrimination value (DV) = 0.57]
- (2) "I can usually handle whatever comes my way" [FL = 0.69; DV = 0.58]
- (3) "I can always manage to solve difficult problems if I try hard enough" [FL = 0.66; DV = 0.48]
- (4) "If I am in trouble, I can usually think of a solution" [FL = 0.56; DV = 0.55]
- (5) "I am confident that I could deal efficiently with unexpected events" [FL = 0.59; DV = 0.49]
- (6) "Thanks to my resourcefulness, I know how to handle unforeseen situations" [FL = 0.49; DV = 0.49]
- (7) "I remain calm when facing difficulties, because I can rely on my coping abilities" [FL = 0.53; DV = 0.45]
- (8) "I can solve most problems if I invest the necessary effort" [FL = 0.46; DV = 0.47]

The reliability coefficients (obtained from the item analysis) for these eight items, based on the responses of 712 individuals, are as follows:

- (a) KR20: 0.80
- (b) KR8: 0.85

As in the case of risk-taking propensity, the validity of the efficacy scale will be tested extensively.

- (a) Internal validity
 - (i) Face validity

It appears that the items of the (self-) efficacy scale presented here deal with a person's ability to solve problems and to cope with difficulties. These abilities are assumed to reflect "efficacy", and therefore the scale seems to have sufficient face validity.

- (ii) Content validity

All eight items were borrowed from the Jerusalem-Schwarzer scale of self-efficacy that has been applied successfully in various cultural settings (see Schwarzer 1998:1). This can be regarded as sufficient evidence of the scale's content validity.

(b) External validity

(i) Concurrent validity

As in the case of the scale for risk-taking ability, it has not been possible to determine the concurrent validity of the efficacy scale.

(ii) Predictive validity

From the literature (De Jong & Fawcett 1981) it can be concluded that efficacy should be an important predictor of whether the respondent has ever considered migrating. Being another element of their category called "personal traits", efficacy (similar to risk-taking propensity) is expected to operate via intermediate variables in its effect on migration intentions.

As in the case of risk-taking propensity it is hypothesised that efficacy would be a good predictor of the probability of migration to have occurred in the past or to occur in future. It is necessary, therefore, to determine the extent to which the suggested efficacy scale succeeds in explaining the following:

- whether or not the respondent has ever migrated voluntarily since reaching the age of 18 years
- the respondent's duration of stay at the current place of residence
- whether or not the respondent intends to migrate in future

A simple Pearson correlation analysis indicated, however, that efficacy was not significantly correlated (at the 5% level) with past migration, duration of stay or migration intentions. A series of regression analyses did provide significant results, however. These are presented below.

(aa) Efficacy and previous voluntary migration

A logistic regression produced the following equation (including social desirability, age and education as control variables, and measured in the same way as before):

$$\text{logit}(p) = -3.02 + 0.018x_1 - 0.009x_2 + 0.064x_3 - 0.302x_4$$

where: p = probability of having ever migrated voluntarily during adulthood
 x_1 = efficacy (expressed as a percentage)
 x_2 = social desirability (expressed as a percentage)
 x_3 = age (in single years)
 x_4 = index of educational attainment

This model correctly classified 63% of the 663 observed cases used in the analysis, and all the coefficients are significant at the 0.01% level.

(bb) Efficacy and duration of stay at the current place of residence

An ordinary least-squares regression (using the same control variables as before) produced the following equation:

$$y = 5.36 - 0.037x_1 - 0.084x_2 + 0.553x_3 + 1.702x_4$$

where: y = duration of stay at current place of residence (years)
 x_1 = efficacy (expressed as a percentage)²⁶
 x_2 = social desirability (expressed as a percentage)
 x_3 = age (in single years)
 x_4 = index of educational attainment

This model explains 21% of the variation in duration of stay, and is significant at the 0.01% level. All coefficients in the model are significant (at the 10% level) except

²⁶ As in the case of risk-taking propensity, the coefficient for efficacy is expected to be negative when modelling duration of stay.

that of efficacy. Efficacy is therefore *not* a significant predictor of duration of residence.

(cc) Efficacy and migration intentions

A logistic regression (using the same control variables as before) produced the following equation:

$$\text{logit}(p) = -3.17 + 0.014x_1 + 0.005x_2 - 0.03x_3 + 0.38x_4$$

where: p = probability of considering to migrate in future
 x_1 = efficacy (expressed as a percentage)
 x_2 = social desirability (expressed as a percentage)
 x_3 = age (in single years)
 x_4 = index of educational attainment

This model correctly classified 64% of the 630 observed cases used in the analysis, and all the coefficients are significant at the 0.01% level.

In all cases but one efficacy proved to be a significant predictor of migration. The one exception is that of period of non-migration (i.e. duration of residence). This should not be surprising, though. Apart from other possibly logical arguments, non-migration is associated with inactivity, and efficacy, being action-oriented, should not be expected to deal effectively with the phenomenon of inertia.

The predictive validity of the suggested efficacy scale has therefore been established sufficiently.

(c) Construct validity

It can safely be said that the efficacy scale proposed here has largely met all three of Trochim's (1999) conditions for construct validity.

MAIN (2001–02) SURVEY

Following the analyses of the initial (preliminary) migration survey, an evaluation of the data from the main migration survey was also undertaken. Confirmatory factor analysis (CFA) was used for the purpose of confirming the findings reported earlier. The findings from these analyses have been reported in detail by Kok and Pietersen (2003) and are therefore merely summarised here.

Item analyses were also undertaken to measure the degree to which particular items discriminate (i.e. differentiate) between different respondents and also estimate the reliability of the items that form part of a particular factor dimension. A series of item analyses was therefore undertaken (on the unweighted data, unless otherwise indicated), again using only those items that survived the confirmatory factor analyses. The findings are described below.

However, before discussing the two suggested scales for risk-taking ability and efficacy, it is necessary to refer to the application of the (shortened) scale for social desirability in this study. This scale did not come out particularly well in the confirmatory factor analyses, but had been applied and tested in various settings in South Africa and should therefore be regarded as sufficiently valid even if not entirely reliable. The six items constituting this scale were measured by means of the same Likert-type five-point scale indicated above.

Social desirability

- (a) "I sometimes try to take revenge, rather than to forgive and forget" [Adjusted discrimination value (DV) = 0.31; Alpha value if this item were to have been deleted, based on the weighted data (α)²⁷ = 0.24]

²⁷ This item was reverse-coded as follows: $N = 6 - O$; where N = new code (value), and O = original code (as it appeared on the questionnaire and the original data set).

²⁷ The value of the Cronbach Coefficient Alpha if the particular item were to have been deleted, using the CORR procedure in SAS with the weighted data. (Please note: The higher the value of α , the less appro-

- (b) "I have sometimes profited unfairly from someone else"* (DV = 0.23; α = 0.28)
- (c) "I am always willing to admit that I have made a mistake" (DV = 0.16; α = 0.31)
- (d) "It does not matter whom I speak to, I am always a good listener" (DV = 0.23; α = 0.29)
- (e) "I am always courteous, even to unpleasant people" (DV = 0.22; α = 0.30)
- (f) "I sometimes feel resentful when I cannot have my own way"* (DV = 0.10; α = 0.43)

Number of cases: 2,872
 KR-20 Reliability: 0.42
 KR-8 Reliability: 0.63
 CCA Reliability²⁸ (on the weighted data): 0.35

It should be noted that the item "I sometimes feel resentful when I cannot have my own way" (DV = 0.10; α = 0.43) performed poorly. (This item must therefore preferably be removed from the social desirability scale during application of the scale in the migration study.) The remaining five items are expected to be sufficient for the purpose of measuring "social desirability" in the study.

Risk-taking ability

- (a) "If the possible reward was very high, I would not hesitate putting my money into a new business, even though it could fail" [Factor loading (FL) = 0.61; DV = 0.44; α = 0.41]
- (b) "I enjoy the challenge of a project irrespective of whether it means a good promotion or the loss of my job" (FL = 0.53; DV = 0.40; α = 0.49)
- (c) "I am the kind of person who likes to take risks" (FL = 0.52; DV = 0.53; α = 0.49)

Number of cases: 2,728
 KR-20 Reliability: 0.59
 KR-8 Reliability: 0.74
 CCA Reliability (on the weighted data): 0.56

These three items performed sufficiently well to warrant their use for measuring "risk-taking ability" in the study.

Efficacy

- (a) "When I am confronted with a problem, I can usually find several solutions" (FL = 0.45; DV = 0.51; α = 0.63)
- (b) "If I am in trouble, I can usually think of a solution" (FL = 0.55; DV = 0.53; α = 0.60)
- (c) "I am confident that I could deal efficiently with unexpected events" (FL = 0.63; DV = 0.52; α = 0.57)
- (d) "I remain calm when facing difficulties, because I can rely on my coping abilities" (FL = 0.52; DV = 0.46; α = 0.60)
- (e) "I can solve most problems if I invest the necessary effort" (FL = 0.46; DV = 0.51; α = 0.62)

Number of cases: 2,808
 KR-20 Reliability: 0.71
 KR-8 Reliability: 0.78
 CCA Reliability (on the weighted data): 0.66

These five items performed well enough to warrant their use for measuring "efficacy".

appropriate the particular item is for that dimension. If α is greater than the overall CCA reliability coefficient it is indicative of a major problem.)

* This item was also reverse-coded.

²⁸ The overall Cronbach Coefficient Alpha (CCA) for the entire dimension, as an index of internal consistency, using the raw (unstandardised) variables.