Journal of Arts & Humanities

Undertaking Effective Cross-Language Questionnaire-Based Survey in Illiterate and Semi-Illiterate Rural Communities in the Developing Regions: Case of Communal Cattle Farmers in Vhembe District of Limpopo Province, South Africa

Mavhungu Abel Mafukata[1]

ABSTRACT

This paper investigated similarities and differences between questionnaire instruments – Source language instrument (untranslated) and target language (translated) instrument employed to gather research data in a cross-cultural background in the developing regions, with special reference to illiterate and semi-illiterate rural communities in the Limpopo Province, South Africa. This paper premises that there are differences of responses between the responses sourced from a source language version questionnaire and that of the translated cross-language version questionnaire. The results of this paper revealed that the majority (60%) of the responses differed in both sets of questionnaires while some (40%) responses on both sets of questionnaires were similar. The implication is that collecting survey data from a cross-language translated version questionnaire might provide a better and reliable data collection option than collecting survey data from a directly translated – or untranslated version during interviews. The differences appear to be influenced mainly by the socio-demographic factors of the respondents. The noted differences might affect the over-all quality of results of the particular survey. This paper argues that such differences might have implications for the design of policy and strategic intervention measures since such interventions might have emanated from flawed results.

Key words: Cross-language, Data collection, Interviews, Language, Translation, *Tshivenda.*
Available Online: 30th November 2014.
MIR Centre for Socio-Economic Research, USA.

[1] University of the Free State, South Africa. Email: drmafukata@gmail.com

1.0 Introduction

Language is fundamental to conducting successful research surveys primarilly looking at the role it plays in research conceptualization and in addition defining differences and commonalities of research concept usage adopted by the researcher (Temple & Edwards, 2002). According to Kavanagh (2000), success in research surveys is also dependent on effective interactive communication amongst the participants. On the one hand, success in communication depends on adequate levels of cultural and linguistic understanding being met. Absence of proper understanding of meanings of words might impede interpretation of words and therefore distort language. Distorted language might lead to miscommunication between the interviewer and the interviewees – resulting in serious consequences on research, planning and implementation of policy processes. It is based on this views that cross-cultural research has been developing into crucial interest of the social sciences – especially in multi-lingual societies of the developing and developed regions respectively. The idea behind cross-cultural research is minimization or total removal of language challenges that might affect quality of research results. Poor language usage – especially that which hinders communication between the respondents and the researcher might be detrimental to the success of research surveys. For example, most empirical research especially in the social sciences in sub-Saharan Africa and elsewhere in the developing regions are conducted in a second or even third language – mostly European languages such as English, French, Spanish and/or Portuguese. These languages are regarded as lingua franca of the region. Temple & Edwards (2002) acknowledge that there has been immense growth of English research especially in non-English speaking societies – or where English is less developed for example – especially in the current dispensation. In most cases, research survey instruments are in original source language, translated or interpreted directly from the source language to the respondents of the research and/or translated from the source to the target language prior to research. Furthermore the mere fact that language in most social science surveys are undertaken in a multilingual setup, might also suggest that linguistic interactions between the survey participants might be constrained by cultural factors such as age, seniority of participants or gender amongst others (Kavanagh, 2000). In other words, something acceptable with this particular age, gender or status group might not be acceptable with the other party. For example, in some cultural groups some words might be sensitive when told to a particular age or gender group than the other.

Research instruments in most of these current studies (Nengovhela, 2010; Nthakheni, 2006; Stroebel, 2004; Randela, 2005) amongst others are in the source language, and it is not clear how these studies dealt with linguistic and cultural factors when conducting their affairs. As a result of the unclear approach as to how the language factor was catered for in these studies, we can only assume that the researchers might have directly translated the questions to the respondents during interviews or interviewed the participants in the source language – which looks highly improbable considering the high illiteracy levels within the communities involved in these studies. If then the interviews were directly translated to the respondents, it is therefore highly probable that the interviewers could not have had the privilege to determine the linguistic correctness and accuracy of the questions before sourcing for responses from the interviewees – let alone to test the cultural acceptability of the language. This approach poses serious data collection challenges especially where the respondents are not conversant enough with the source language used for the interviews. Poor language proficiency and amongst others sensitivity of the language among the respondents might lead to misunderstanding and unwillingness to answer the questions and therefore lead to collection of distorted and misleading data. This response might eventually affect the quality of the results and the subsequent findings of the research survey. On the one hand, since the interviews are conducted by different data collectors, it might not be possible for all the data collectors to translate all the questions to the respondents the same all the time. In this case, respondents are likely to understand the questions differently – and therefore impacting on the reliability of the data collected.

However, we consider the fact that most researchers who adopt direct translation of research instruments approach argue that the research data collectors are trained prior to interviews to minimize envisaged challenges such as these. I argue in this paper that such trainings may never be

adequate enough to remove all possible cultural challenges of language among the respondents, remove the obvious data distortions and in addition, curb the huge possibilities of producing flawed and unreliable research results at the end. In fact, this view is corroborated by Su & Parham (2002) who contend that language inaccuracies at times resulting from linguistic and cultural differences between the source and target languages in the translation of research instruments are common in survey research. These authors furthermore argued that such inaccuracies might compromise the quality of the survey data and subsequently the results. Since most research is carried out for aiding policy and strategic development, such data might lead to flawed research reports which might affect policy and strategic directions. In order to remove the probability of flawed research reports which might implicate on policy and strategy, it is therefore crucial to consider the recommendations by Harkness & Schoua-Glusberg (1998) who presented that a common interpretation and analysis of a particular research survey might best be achieved if the data was collected from the same instrument. The suggestions made by these authors advocate for survey researchers to do away with the approach where data collectors translate questions to the respondents directly from the questionnaires. In other words, Harkness & Schoua-Glusberg (1998) argue that translating source questionnaires into the language of the interviews prior to the surveys, for example, might assist in creating equivalent research instruments.

Over the years, various researchers have adopted and preferred different ways of treating language issues in research. For example, Nengovhela (2010), Nthakheni (2006), Stroebel (2004) and Randela (2005) among others were not clear on how they collected data from questionnaire instruments that were not in the languages spoken by the respondents in their studies. On the one hand, Hangara (2011), Juma (2009) and Nanja (2010) defined and elaborated on how they dealt with language issues in their surveys. It must be said though that Temple & Edwards (2002) concede that very few researchers spell out in their research methodologies the effects of translation of language in research. However, it might be that these studies (Nengovhela, 2010); Nthakheni, 2006; Stroebel, 2004; Randela, 2005) adopted direct translations from the source language research instrument versions to the target language of the respondents through bilingual agents conducting their data collection. The choice of this approach in a way underestimates potential inaccuracies and compromised data quality which might emanate from the process. The limitation in this regard is that direct translations might not provide the researcher with the luxury of comparison of the accuracy and equivalence of certain concepts prior to data collection (Su & Parham, 2002).

On the other hand, Hangara (2011), Juma (2009) and Nanja (2010) adopt a different approach to Nengovhela (2010), Nthakheni (2006), Stroebel (2004) and Randela (2005) in that they spell out in their research methodologies the effects of translation of language in data collection by putting emphasis on the need to translate their source questions to the target language of the respondents. However, these studies also have nevertheless adapted to directly translating the questionnaires without satisfying themselves with the correctness or accuracy of the translations. There are however different reasons as to why researchers recommend questionnaire translations. For example, Hangara (2011) first translated the English Version questionnaire into the local *OtjiHerero* to collect the data for a study conducted in Omaheke, Namibia – because the majority of the respondents there spoke OtjiHerero while on the other hand Juma (2009) realised the need to translate Kiswahili which is the national language of Kenya when collecting data for a study among the so-called Turkana nomadic farmers of Kenya because only a few of the respondents were conversant in *Kiswahili*. Instead, data was collected through Ngaturkana; a local language spoken by the Turkana nomadic cattle farmers among whom the study was conducted. Although the researcher could not necessarily speak Ngaturkana, there was an interpreter who translated the questionnaire into Ngaturkana to the respondents during interviews. The similarities of *Ngaturkana* and the so-called *Dholuo*; being the mother tongue of the researcher assisted the researcher to understand what was transpiring during the process. Nanja (2010) also resorted to the local vernacular in collecting data among the *Tonga* of Monze District in Zambia while collecting data for a study in the Nkabika, Malomo and Bulimo villages of Mujika. The data was collected in the local Tonga and later translated into English. Although it is not clearly indicated in the

Nanja's study, it is deducible that Nanja (2010) might have translated the survey instrument from English into Tonga prior to the interviews. However, this assertion remains speculation.

Some researchers argued that research data gathered in a second language other than the language spoken by the respondent stands to be vulnerable to ambiguity, poor quality and misinformation (Douglas & Craig, 2007; Jiang et al., 2006). These are those who argue that it becomes imperative of the researcher to translate the survey questionnaires from any other language not spoken by the respondents to the language of the respondents. In other words, the questions are read out to the respondents in the language of the interview – not translated directly during interviews to the respondents. Translations prior to data collection minimize most challenges because the researcher will have had the luxury to compare and adjust any differences in the languages of the survey instruments. Clearly, translation of the survey instrument to the language well understood by both the respondent and the researcher is critical for clear understanding of the responses given by both parties (Peil, 1995) and further to that to avoid any possible ambiguity (Kyrillidou et al., 2003; May, 1997; Peil, 1995). This is critical especially among most African communities who are expected to struggle with survey instruments in a foreign source language such as English as a result of their poor literacy levels and language command of most foreign European languages (Akpalu, 2005). In the case of this study area, the respondents were expected to have low literacy levels and poor multilingual background especially with regard to English - which might be constraining to them answering questions from an English questionnaire.

The objective of this paper is to show that there are differences in responses between the answers given by respondents on an interview from a direct translation from a source language version questionnaire and the responses obtained from a cross-language translated version of the target language questionnaire. Furthermore, this paper argues that such differences have an effect on the results and final report of the particular study – since the results were bound to be different. The paper further argues that such differences might also affect policy and strategic considerations as such policy and strategy might have been emanating from flawed report. The paper finally attempts to establish the effect of selected demographic characteristics of the respondents on the quality of the data – however focusing on age and educational levels of the respondents.

2.0 Description of Study Area and Methodology

2.01 The Study Area

This paper was undertaken in the Musekwa Valley of the Vhembe District in Limpopo Province, South Africa, South Africa. There are eight villages; Afton, Dolidoli, Khomele, Maangani, Maranikhwe, Musekwa, Sane and Strathaird comprising the Musekwa Valley study area. These villages are under the leadership of traditional chiefs known as *Vhamusanda* or *Khosi*. The dominant economic practice of this study area is livestock farming – especially chicken, donkeys, goats, sheep and cattle. The largest livestock population is that of goats and cattle. In addition to livestock, some of these farmers also practice crop farming. In other words, a mixed livestock-crop farming system is dominant in the area. The dominant language spoken in this study area is Tshivenda which is both the spoken and formal education language in the area. However, there are some residents of this area whose home language is *Northern Sotho* – especially at Khomele and Strathaird villages where there is a combination of Tshivenda and Northern Sotho languages. However, *Northern Sotho* is not part of the formal education system of the area. The majority of the population in these eight villages are women, old and of poor literacy levels. Most households are headed by unemployed men who rely on livestock farming – especially cattle and various government grants for livelihood.

2.02 Methodology

Six (n=6) randomly-selected participants were drawn from the fifty five (n=55) communal cattle farmers used for the main doctoral study. The selected six participants were also used during the

piloting and pre-testing of the questionnaire instrument used for the main study. Only "Section C" of the questionnaire instrument used for the main study was selected for use to gather data for this paper. This section was chosen for its limited number of questions (n=10) which was envisaged to be more convenient and less complicated for the purpose of this paper. Fewer questions ensured a simplified process of data gathering, entry and analysis – also saving on time and resources for the researcher. Since the cross-language translations had already been handled during the process of the main doctoral study, what the process of this paper did was to interview the respondents using the two sets of questionnaire instruments – the English and the *Tshivenda* Versions. However, the guidelines followed in the translation of the source language to the target language are provided in this paper to refresh the reader. As indicated here-under, the guidelines of translating questionnaires as suggested by Douglas & Craig (2007) and ICN (2005) were adopted:

- ***Pragmatic translation***: Ensures that the content of the source language used in the questionnaire (English) is accurately communicated in the target language of the respondents (Tshivenda)
- ***Aesthetic-poetic translation***: Ensures that the moods and feelings of the source language used in the original questionnaire are maintained in the target language
- ***Ethnographic translation***: Ensures that meaning and cultural content of the translated local dialect questionnaire matched the expectations of the respondents, and
- ***Linguistic translation***: Ensures that the grammatical forms and equivalent meanings in the target language of the respondents were correctly presented in the local *Tshivenda* dialect questionnaire.

After the translations, a comparison of the source language questionnaire instrument and the back-translated version was conducted by a language specialist who was not involved in the original translation of the questionnaire from English to *Tshivenda*. This was done in order to detect and correct any possibility of dissimilarities between the questionnaires and to achieve high quality measurement of the questionnaire. Naotaka (1992) submitted that in order to guarantee quality of the translated versions, the back-translation technique is necessary to assist in finding a degree of congruence between the original version and back-translated transcripts. In addition, the translations were to ensure that the final version of the questionnaire was both relevant and effective (Bailey, 1987; Goddard & Melville, 2001). A two-step data collection approach was used. First, the six participants were interviewed on a one-on-one approach in their homes using direct translation from the source language – and the responses were then recorded for analysis. Following this was the interviews using the target language questionnaire instrument already translated in terms of the guidelines presented in this paper. Most crucially is the fact that the interviews were conducted in two separate meetings, different days and enumerators. This allowed for the respondents to forget the responses they had given in the previous interviews which could have compromised the quality of the data. The results of the two-step interviews were thereafter analysed, and the results reported as indicated in Table 1.

3.0 Results and Discussion

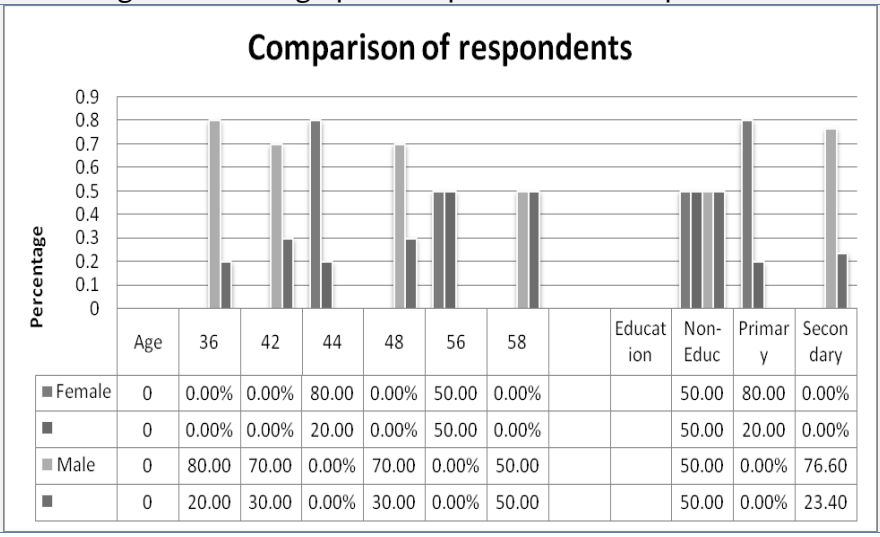
As indicated in Table 1, the results of this paper revealed that there were indeed differences and similarities in the responses given by the communal cattle farmers during data collection between the English and Tshivenda Versions of the questionnaires. The results revealed that six out of ten (60.0%) of the questions had different responses while four out of ten (40.0%) had similar responses on both sets of the questionnaires. Differences and similarities of 50.0% were recorded in questions three, five and nine of the questionnaires. The greatest differences were recorded in questions four (16.7%; 83.3%), seven (33.3%; 66.7%) and eight (66.7%; 33.3%) of the English and *Tshivenda* Versions of the questionnaire respectively. For example, five pairs (50.0%) of the questions were similar and five pairs (50.0%) differed for the first male respondent. For the second male respondent, eight (80.0%) and two (20.0%), third male respondent, six (60.0%) and four (40.0%), and for the fourth male respondent, seven (70.0%) and three (30.0%) emerged from the analysis.

Table 1: Comparison of responses given by communal cattle farmers on the Venda and English questionnaires

			Socio-Demographic characteristics			Question																			
						1		2		3		4		5		6		7		8		9		10	
			Gender	Age	Educa-tion	English	Venda	English	Venda	English	Venda	English	Venda	Engli	Venc	Engli	Venc	Engli	Venc	Engli	Venc	Engli	Venc	Engli	Vend
Respondent	1		M	58	N	1	1	1	1	1	2	1	3	3	4	1	1	1	3	1	1	2	4	1	1
	2		M	42	S	1	1	1	1	1	1	2	4	1	1	1	1	3	1	1	1	1	2	1	1
	3		F	56	N	1	1	1	1	2	1	4	2	2	3	1	1	5	4	2	3	3	5	4	4
	4		M	48	S	1	1	1	1	2	2	4	2	3	3	1	1	1	3	3	2	4	4	5	5
	5		M	36	S	1	1	1	1	1	2	1	3	4	4	1	1	1	1	2	2	4	4	3	3
	6		F	44	P	1	1	1	1	2	2	2	2	2	3	1	1	2	2	2	2	1	2	1	1
		Status				Similar	Differer	Similar	Differe	Similar	Differe	Simila	Differe	Simil	Differ	Simil	Differ	Simil	Differ	Simil	Differ	Simil	Differ	Simil	Differe
		Total				6	0	6	0	3	3	1	5	3	3	6	0	2	4	4	2	2	4	6	0
		Percentage				100	0	100	0	50	50	16.7	83.3	50	50	100	0	33.3	66.7	66.7	33.3	33.3	66.7	100	0

Evidently, all the male respondents would not equally match all the responses correctly. The implication is that all the male respondents struggled to understand the questions sufficiently – and hence the different responses to the same questions. On the other hand, all the female respondents also wouldn't equally match the responses correctly. The results revealed that four (40.0%) of the questions of the first respondent were similarly matched while six (60.0%) differed. The second female respondent matched eight (80.0%) of the questions similarly while two (20.0%) were differently matched. Comparatively, all the sets of respondents; the male and female would not match all the questions similarly. Due to the differences in the responses, it means that the research reports emanating from these two sets of data will bear significant differences. Language is playing a major role in determining the final research outcome. It is also clear that the results of this paper revealed that responses differed on age, gender and educational levels of the respondents as indicated in figure 1.

Figure 01: Demographic comparison of the respondents



	Age	36	42	44	48	56	58		Education	Non-Educ	Primary	Secondary
Female	0	0.00%	0.00%	80.00	0.00%	50.00	0.00%			50.00	80.00	0.00%
	0	0.00%	0.00%	20.00	0.00%	50.00	0.00%			50.00	20.00	0.00%
Male	0	80.00	70.00	0.00%	70.00	0.00%	50.00			50.00	0.00%	76.60
	0	20.00	30.00	0.00%	30.00	0.00%	50.00			50.00	0.00%	23.40

Males who have attained some formal educational level were able to match the majority of their questions and responses similarly (60.0% to 80.0%) while those with lesser education matched lesser responses similarly (50.0%). This might suggest that the level of education among the respondents had effect on the nature of their response to questions as expected that proficiency of language might have a link with educational levels of the respondents – especially in rural South Africa (Akpalu, 2005). Apparently, the differences in the responses to the questions might further explain the fact that the respondents might have lacked proper understanding of the questions due to linguistic short-comings (Peil, 1995); either as ambiguity or mere misunderstanding (Douglas & Craig, 2007; Jiang et al., 2006) – which should be avoided in data collection (Kyrillidou et al., 2003; May, 1997). On the other hand, female respondents with some formal educational levels managed to match 80.0% of their responses similarly while those with lesser formal education levels were able to match only 50.0% of their responses similarly. From the results of this paper, it is evident that response patterns were also largely influenced by the age of the respondent. For example, older respondents (56 to 58 years) matched lesser responses similarly (40.0% to 50.0%) while younger respondents (36 to 48 years) matched the majority (70.0% to 80.0%) of their responses similarly. The response patterns might suggest that language differences affected the older respondents more than the younger respondents – and furthermore, the response group that experiences language challenges was therefore affected in its response.

4.0 Summary and Conclusion

Cross-language data collection remains a critical data collection approach in the developing regions – especially where research has to be conducted in a language other than the locally spoken dialect. This paper demonstrated that directly translated data collection approaches produce different outcomes to

instruments of the locally spoken languages – suggesting that efforts should be made to collect data from the target language research instrument. In addition, demographic characteristics of the respondents of research such as age and educational level among others play a major role in determining data quality as the data is highly dependent on these characteristics. Data from older and less educated respondents should be collected from simplified research instruments – especially with regard to language. Collecting data from a directly translated research instrument has proven to impact on the quality and reliability of the data because older and less educated respondents have shown poor response to that kind of research instruments. In short, I suggest that it be mandatory that data collection from respondents who are second, third or so on speakers be conducted from translated survey instruments so that data gatherers should read the questions directly to the respondents – not directly translate from the questionnaire to the respondents.

References

Akpalu, D.A. (2005). *Response scenarios of households to drought-driven foot shortage in semi-arid areas in South Africa*. Unpublished MA thesis. University of the Witwatersrand, Johannesburg.

Bailey, K.D. (1987). *Methods of Social Research*. London: Collier MacMillan.

Douglas, S.P. & Craig, C.S. (2007). Collaborative and interactive translation: An Alternative Approach to Back Translation. *Journal of International Marketing*, American Marketing Association 15(1): 30-43.

Goddard, W. & Melville, S. (1996). *Research methodology: An introduction.* Lansdowne: Juta.

Hangara, G.N., Teweldemedhin, M.Y. & Groenewald, I.B. 2011. Measuring factors that can influence cattle supply response to the market in Namibia: Case study from Omaheke communal farmers. *Journal of Agricultural Extension and Rural Development* 3(8): 141-146.

Harkness, J.A. & Schoua-Glusberg, A. (1998). Questionnaires in translation. ZUMA – Nachrichten Spezial: 87-126.

Harzing, A. (2006). Response styles in cross-national surveys research: A 26 country study. *International Journal of Cross-cultural Management* 6(2): 1-38.

ICN. (2005). Translation Guidelines for ICNP. Geneva, Switzerland.

Jiang, S., Lambert, E.G. & Wang, J. (2006). Issues in cross-cultural justice survey research: A case of U.S. and China collaboration. Internet Journal of Criminology. [online] Retrieved from: http://www.internetjournalofcriminology.co. [Accessed 23 June 2009].

Juma, R.O. (2009). *Turkana livelihood strategies and adaptation to drought in Kenya.* Unpublished Ph.D thesis. Victoria University of Wellington. Wellington, Australia.

Kavanagh, K. (2000). Words in a Cultural Context. Paper presented as a keynote address at the Fifth International Conference of the African Association for Lexicography, University of Stellenbosch, Stellenbosch, 3-5 July 2000. Lexikos 10 (AFRILEX – reeks/series 10:2000):99-118.

Kyrillidou, M., Olshen, T., Heath, F., Bonnely, C. & Cote, J. (2003). Cross-cultural implementation of LibQUAL+TM: the French Language experience. [online] Retrieved from: http://www.libgual.org.documents/admin/Northumbria%20final6.pdf.[Accessed 24 May 2008].

May, T. (1997). *Social Research issues, methods and process*. Buckingham, United Kingdom: Open University.

Nanja, D.H. (2010). *Dissemination of climatic information to small-holder farmers: A case study for Mujika area, Zambia*. Unpublished Ph.D thesis. University of the Free State, Bloemfontein.

Naotaka, W. (1992). Item Response Theory and Language Translation. An Aid to Comparative Studies of Management in the United States and Japan. [online] Retrieved from: <http://www.nanzanu.ac.jp/AMERICA/kanko/14-2watanabe.pdf.pdf. [Accessed> 21 May 2008].

Nengovhela, N.B. (2010). *Improving the wellbeing of people dependent on the low-income beef industry in South Africa*. Unpublished Ph.D thesis. University of Queensland, Australia.

Nthakheni, N.D. (2006). *A livestock production systems study among resource-poor livestock owners in the Vhembe District of Limpopo Province*. Unpublished Ph.D thesis. University of the Free State, Bloemfontein.

Peil, M. (1995). *Social Science Research Methods, a handbook for Africa.* Nairobi: East African Educational.

Randela, R. (2005). *Integration of emerging cotton farmers into the commercial agricultural economy.* Unpublished Ph.D thesis. University of the Free State, Bloemfontein.

Sherry, E. (2002). *Constructing partnerships: A Delphi study of shared resources management in the North Yukon*. Unpublished Ph.D thesis. University of Northern British Columbia.

Stroebel, A. (2004). *Socio-economic complexities of small-holder resource-poor ruminant livestock production systems in sub-Saharan Africa*. Unpublished Ph.D Thesis. University of the Free State, Bloemfontein.

Su, C.T. & Parham, L.D. (2002). Case Report – Generating a valid questionnaire translation for cross-cultural use. *American Journal of Occupational Therapy* (56): 581-585.

Temple, B. & Edwards, R. (2002). Interpreters/translators and cross-language research: Reflexivity and border crossings. *International Journal of Qualitative Methods* 1(2).