This is the published version (version of record) of:


Available from Deakin Research Online:

http://hdl.handle.net/10536/DRO/DU:30032196

Reproduced with kind permission of the copyright owner.

Copyright : © 2010, ANZAM
Minimising Total Survey Error when Measuring Entrepreneurship in Tonga

Dr. Howard Frederick

Centre for Entrepreneurship, Innovation & Community, School of Marketing and Management,

Deakin University, Melbourne, Australia

howard.frederick@deakin.edu.au
ABSTRACT

This study illustrates how the Total Survey Error (TSE) paradigm can identify and help reduce multiple sources of error inherent in survey work in the developing world. Of particular concern are mode errors and coverage errors caused by the 'theoretical teledensity threshold' of doing phone surveys in developing countries. The study outlines ways to improve response rate and to avoid interviewer and measurement error. It narrates the sampling design and its limitations as well as some of the qualitative aspects of total survey quality such as, translation, ethics and budgeting. The final section discusses implications for further research in statistical auto-correlation and data gathering using PDAs.

Keywords: Research methods – surveys, international culture, entrepreneurship, indigenous entrepreneurship, methods in entrepreneurship research.

INTRODUCTION

Survey quality is closely related to survey measurement, and monitoring and minimizing of survey error (Biemer et al., 2004, Biemer and Lyberg, 2003, Groves, 2004a, Lessler and Kalsbeek, 1992, Lyberg, 1997). The Total Survey Error (TSE) paradigm has emerged as a way to examine the multiple sources of error that arise from the various components of survey instrumentation (Weisberg, 2005, Groves, 2004a, Groves, 2004b). Although it has become widespread among professional researchers in the survey field, to date TSE has received scant attention in the field of entrepreneurship research or business research in general. In particular, the Global Entrepreneurship Monitor (GEM) research, which this paper addresses, has not properly addressed Total Survey Error as it continues to explore entrepreneurial activity developing in countries. This study uses the example of a household survey incorporating in-person interviews of business and social entrepreneurship in the Kingdom of Tonga to illustrate these issues. This article seeks to show that the TSE paradigm has relevance across all these dimensions in the context of the cross-cultural implications entrepreneurship research in the poorest countries.

SOURCES OF ERROR IN GLOBAL ENTREPRENEURSHIP SURVEY RESEARCH

In this paper I am going to examine two supersets of error in entrepreneurship surveys, one is administrative and practical; the other is cultural and practical.

Administrative and practical sources of error
The TSE approach examines frequent sources of error in survey research:

*Sampling error* is the random variation resulting from a random sample. It is caused by observing a sample instead of the whole population, e.g. the underrepresentation of Maori New Zealanders as owners of fixed-line telephones. It occurs when a sample does not accurately represent the larger population of interest.

*Coverage error* means that the sampling frame does not match the population, e.g. carrying out fixed telephone line surveys in low a teledensity country where fixed-line usage is low and cellphone usage is high.

*Non-response errors* occurs when answers are missing due either to ‘unit non-response’ (failure of an individual to respond) or ‘item non-response’ (failure to answer a particular item), e.g. due to sensitivities on the household income question or a ‘don’t know’ response to avoid embarrassment, respectively.

*Measurement error* occurs when the survey instrument does not accurately measure what was to be measured, due often to poor question wording, interviewer carelessness, or respondents not being appropriately contextualized. While every attempt is made to avoid the second two, after ten years of GEM studies we are still not certain that ‘TEA’ is actually measuring what we want to measure.

One important component of *measurement error* that has important consequences for GEM is “mode effects”, which refers to differences that occur among phone samples, household surveys, in-person or face-to-face interviews, or Internet surveys using the same questionnaire. (This study in its conclusion recommends an experiment to help overcome this problem.)

*Post-survey error* is error due to scanning of forms, programming errors, or incorrect weighting.

While most GEM countries do not experience this when they use CATI surveys, in developing countries numerous problems of a practical nature arise in this regard.
Cultural sources of error and practical problems in the developing world

Beyond administrative and practical sources of error in survey research is the cultural component. Weisberg (2005) cautions that 'simply applying the Western mode of surveys to non-Western societies may not produce meaningful results'. There is a variety of potential sources of error in survey research that relate to culture. Some of them include:

*Cognitive differences.* Respondents vary culturally in cognitive tasks such as question interpretation, information retrieval from memory, judgment formation, and response editing.

*Equivalence.* More than equal selection probability of respondents, equal co-operation rates and equivalent coding schemas, the researcher must take into account language and idiom.

*Etic and emic nature of responses.* Etic questions elicit responses that can be compared to other cultures whereas emic questions elicit responses that are meaningful to one culture alone (Pike, 1967).


*Response styles.* What is called ‘extreme response bias’ (the tendency to select the end-point of a scale) has been shown to vary culturally (Arce-Ferrer, 2006, Bachman and O'Malley, 1984, Clarke III, de Jong et al., 2008, Harzing, 2006, Stening and Everett, 1984, van Herk et al., 2004, Warnecke et al., 1997). The counterpoint to this is ‘acquiescence bias’ (the tendency to agree with questions regardless of content) and this differs among cultural groups (Aday et al., 1980, Cunningham et al., 1977, Hui and Triandis, 1989, Javeline, 1999).

*Socially desirable reporting.* Socially desirable reporting is a systematic tendency to give answers that make the respondent look good (Steenkamp et al., 2010) and this has the potential to introduce response bias (Johnson and van de Vijver, 2002).

*Translation error.* Questionnaire translation is a transfer process from a source language into an ‘optimally equivalent’ expression in a target language. Equivalent meaning must be transmitted,
but not necessarily through literal translation. An iterative process of translation and back-translation is often employed.

_Sheer physical difficulty._ Low response rates, research administration burdens, lack of co-ordination, the time-consuming nature of such endeavors, and differences in cultural perspectives of the researchers, not to mention extreme weather conditions, are some of the major challenges of international survey research (Easterby-Smith and Malina, 1999, Harzing, 2001, Martinez, 1987).

In this paper, an actual case study applies the Total Survey Error (TSE) paradigm to the assessment of entrepreneurial activity in the Kingdom of Tonga. The challenge to this study’s researchers, enumerators, translators, and administrators was to design a survey that reduced as much error from multiple sources.

**THE KINGDOM OF TONGA**

The Kingdom of Tonga comprises 169 islands, 36 inhabited, and stretches over a distance of 800 kilometers from north to south. Tongans has a high life expectancy rate (70.7 years) and a high literacy rate (98.9%). Although Tonga is poor, its GDP per capita ($4,600) puts it above the category of ‘least developed countries’ (CIA, 2010).

In terms of cultural characteristics, high value is placed upon collective ownership of community land and on sharing of income, assets and activity. Tongan social structure is based on collective rather than individualistic values, which are secured through deep-rooted relationships at multiple levels. Social order and material comfort are maintained through protocols and values of tradition enforced consensually within a context of family, community and society (Bain, 1993, Campbell, 2001, Duckitt and Parra, 2004, Kaeppler, 1999, Moore et al., 2005, Morton, 1996, Prescott, 2009, CIA, 2010). Non-compliance is communal norms ‘is an invitation to anger, disrespect and ostracism’ (Rao, 2005).

_The GEM model_

Government ministries and development aid agencies are increasingly motivated to fund surveys of entrepreneurial activity in developing countries because of the connection between entrepreneurship and economic development (Wennekers and Thurik, Carree and Thurik, 1999, Audretsch et al., 2002).
This study examines a household survey of the Kingdom of Tonga as part of the Global Entrepreneurship Monitor (GEM), a ten-year-old cross-national and cross-cultural benchmarking tool for the measurement of entrepreneurial activity. GEM defines entrepreneurship as any attempt at new business or social enterprise by an individual, teams of individuals, or established businesses or social enterprises. GEM has three objectives:

- To measure differences in the level of social and business entrepreneurship activity between countries
- To uncover factors determining levels of entrepreneurial activity
- To identify policies that may enhance levels of social and business entrepreneurial activity.

GEM uses an adult population survey normally of 2000 respondents per country to identify the proportion of the population who are start-up business entrepreneurs, established business owner/managers, start-up social entrepreneurs, or established owner/managers of a social enterprise.

**SOURCES OF ERROR IN GEM RESEARCH**

**Mode and coverage error**

Many researchers view household surveys as having less error than other types of surveys because of higher response rates and greater rapport of interviewers with the respondent. But they are much more costly than other modes of surveys, take much longer to complete in the field, and have complex post-survey and data-entry procedures. They are also more likely to be affected by the cultural issues surrounding face-to-face communication.

The first issue was ‘teledensity’, or proportion of the population that have fixed line telephone service. Like most researchers, GEM teams prefer to use computer-assisted telephone interviews (CATI) of fixed-line telephones, mobile telephones (or both) in order to have shorter field times, better interviewer control, and most importantly lower cost (since funding in each country must be individually raised within the academic environment) (Frederick et al., 2010). But CATI in particular surveys have many shortcomings, the most troublesome of which is coverage error. *Coverage error*
results from selecting only a portion of the population (e.g. those with fixed-line telephones) for participation in the survey.

The situation is compounded these days by the widespread use of cellphones (Brick et al., 2006). Entrepreneurs in developing countries are high users of cellphones (Donner, 2006). Weighting techniques may sometimes help overcome these deficiencies, but with varying degrees of success (Keeter, 2007).

In effect, telephone surveys in developing countries confront a coverage error called the ‘theoretical teledensity threshold’ below which measurement is not reliable. Where this threshold may lie is a matter of discussion. Obviously, at the low end, ‘telephone interviews are clearly not an option ... where teledensity is under 10%’ (Fu and Chu, 2007). Kempf and Remington (2007) agree, ‘If the rate of telephone coverage is extremely low—as is the case in many developing countries—a telephone survey will likely be impractical and inefficient’. As Bernard (1995) cautions, ‘In the Third World, telephone surveys are out of the question, except for some urban centers, and then only if your study requires a sample of relatively well-off people’. Tonga has a teledensity rate of 25%, comparatively high by developing world standards measured by the International Telecommunication Union (ITU, 2010). However, the Government Statistician of the Kingdom of Tonga, the subject of the present study, indicated that he would not recommend using a telephone survey in his country because of unrepresentedness due to low teledensity of 25% (Finau, 2009).

Response rate

Response rates of household surveys in developing countries are often higher, sometimes much higher, than in the developed world. In a culture widely known for its preference for interpersonal interaction, and higher observed need for uncertainty avoidance and ambiguity reduction, Tongans were more likely to prefer the face-to-face modes of information exchange and were less likely to cooperate with those modes with less social interaction. From a cultural perspective, it was essential to have government and church support. Publication via TV, radio and newspapers legitimated the survey in the eyes of the population. With resonance in the public sphere, most people were likely to cooperate. In the end, we achieved over 90% response rate.
Interviewer error

While Tonga was the smallest country in the GEM 2009 cycle, it required one of the largest teams. Needless to say, the quality of such a household survey incorporating face-to-face interviews depends on the quality of its enumerators. Recruits had to be people who had flexible time schedules or who were not currently holding jobs. They also had to be willing to spend time away from home. Beyond this, all had to be culturally grounded in their society. All underwent a rigorous four-day training and completed at least ten practice interviews in the field that were critiqued by the lead researchers.

Training manuals explained the purpose of the survey, basic tasks to be performed by staff, procedures for enumeration of households, finding and selecting their respondents, techniques for recording the responses, methods for obtaining compliance of respondents, not to mention how to protect the surveys from inclement weather and protect oneself from unfriendly animals. After each test cycle, the entire group engaged in a procedural and cultural debriefing.

Measurement error

Considerable attention was paid to the construction of the questionnaire. The GEM organization supplied an English-language version of questionnaire as well as a question-by-question commentary to assist the translators. The questionnaire had been validated cross-nationally for ten years and is annually discussed and updated at the GEM consortium meetings.

The first section of the questionnaire contained twelve questions asked of all respondents that were easy to answer and pertained to non-sensitive topics. Sensitive questions such as income level and ethnicity were put at the end. At the end of the questionnaire, we required a respondent ID as well as data on gender, age in years, age in categories, educational attainment, employment status, annual household income, household size, region, city, and strata indicator. We also added ethnicity, whether the person had lived abroad, and, if so, how long had they been back in Tonga.

SURVEYING ENTREPRENEURSHIP IN TONGA

Questionnaire preparation
The GEM organization supplied an English-language version of the Adult Population Survey (APS) questionnaire as well as a question-by-question commentary to assist the translators. The questionnaire was prepared in English and Tongan; the enumerators were fluent in both languages. Many terms in the English APS that do not have an equivalent in the Tongan language. Therefore it was important for the team to find the correct language to use to interpret the many concepts within the APS. A useful reference for this process was (Harkness et al., 2003). The survey was first translated by an eminent language expert and then back-translated by the Ministry of Education’s Tongan Language Board Examiner. Beyond that, the survey was reviewed top to bottom by the Tongan Government Statistician. Finally, our 38 bilingual enumerators went over the translation with a fine-toothed comb over four days of training.

The GEM household survey questionnaire was composed of several sections. Based upon the answers to the filter questions, a respondent would proceed to one or more sections for: start-up business entrepreneurs, start-up social entrepreneurs, business owner-managers, social enterprise owner-managers, business angels, and business exiters.

Survey design

Within in the cultural context of the Kingdom of Tonga, we chose to carry out a multi-stage, stratified, area probability sampling technique that utilized a sampling frame based on detailed enumeration maps. The core design was greatly facilitated by the Tongan Department of Statistics, which provided size information and location of census blocks throughout the country, and maps of those census blocks selected by probability proportional to size sampling within strata.

Sample selection was carried out within strata. Each stratum was sampled as an independent sub-population, out of which clusters, then households and individuals were randomly selected. Second-stage units within each selected sampling unit were the enumerated households as officially defined. Third-stage units were the people interviewed.

To maximize representativeness, we used a two-step process beginning with the procedure proposed by Kish (1965) for face-to-face interviews that first requires listing the age and gender of all
household members; follow by a procedure by Salmon and Nichols (1983) for selecting individuals on the basis of which adult has had the last (or next) birthday. In the end, within each stratum, about one in seventeen of the households, and two in every seven census blocks, were chosen. We then set the start number and a fixed skip interval determined as the number of households divided by 12, since the sample size within clusters was twelve households and one person per selected household.

Next came the issue of sample size. The major issues confronting surveys are always the trade-offs in precision (reliability) of the survey estimates, quality of the data collected, and cost in time and money of data collection, processing and dissemination. The optimal design strategy had to be carefully considered against the increased complexities in management, costs and the effect on non-sampling errors, as well as the added cost of traveling to thirty-six inhabited islands.

The population of the Kingdom of Tonga is very small, estimated at 120,898 in 2009 (CIA, 2010). Normally the GEM project requires a sample size of at least 2,000 respondents, which would have meant interviewing almost 2% of the eligible population. Because Tonga is made up of many small islands, costs of surveying are high, so the required accuracy depended as much on efficiency of design, via careful stratification and sample allocation to strata, as on sample size \textit{per se}. Given the high cost of plane flights and the precarious nature of ship transportation, the sample size was budgeted at 1,200. In the end, we collected 1,184 completed surveys.

The next step was to devise the sampling scheme that would be appropriate to the culture, topography and family structure of Tonga. Before sample selection, each primary sampling unit was assigned a measure of size based on the number of people or households recorded for it during the recent census and re-enumeration. Once the sample selection was completed, we carried out a procedure to list all households in each selected primary sampling unit. The objective was to create an up-to-date sampling frame from which households could be selected. After re-enumeration, one house was randomly chosen to start, and then a systematic sample of the remaining houses was made by skipping designated houses in the numbering sequence, and sampling the rest. This meant identifying a household (a group of people who normally eat together) and, after finding someone at home, the researcher first asked how many people aged 18 or over there were in the household. Then she asked
who had the next birthday. In Tongan society it is common for an 18-year-old who answers the door, and whose next birthday it is, to defer to an elder. However, the objective was to interview only this selected person and to substitute no one else under any circumstances. This did lead occasionally to the selected youth answering questions in the presence of an elder.

The challenge was always to achieve a balance between cost savings and loss in precision through the selection of clusters. In the case of Tonga, the budget managers were worried that we would have to send researchers to the farthest-flung island. We finally decided that remote volcanic rim island of Niuafo'ou would be included in the sample and sent one researcher there on a once-a-week flight to collect about 25 responses from the 650 residents.

Field work

To assure each step in our effort to minimize total survey error, we used teams of three researchers in the fieldwork. Having teams allowed us to achieve a higher level of supervision and (most important in Tonga) to co-ordinate the means of transportation for the interviewers. The quality of work done by interviewers was assessed by a system of fieldwork supervision. Each supervisor was responsible for a small number of interviewers.

Field-testing the questionnaire was carried out in stages. The first stage involved sending the researchers out into the street, in the neighborhood of the training facility, to test selected sections of the questionnaire on 10-15 individuals. Trainees then returned to the seminar room and detailed discussions followed on formatting, translation, and positioning of the questions in the survey as well as the cultural aspect of how the survey was being received.

The second stage was a comprehensive pilot test. This was a larger operation, involving about 100 households from the sample. We chose census blocks within easy travel of the training site, and conducted the pilot in both urban and rural areas. All members of the team participated in the pilot test and observed each other as much as possible.

Data entry used a laptop computer. Using laptop computers in the field is not without its difficulties. Problems can include unreliable electricity, dust, heat and high humidity. We used Apian's
SurveyPro™ software to design the forms and then Remark Office OMR™ form-processing software using a Fujitsu duplex image scanner, all powered by a laptop (2gb RAM) with back-up batteries. Remark recognizes optical marks (bubbles and checkboxes), machine-generated characters (OCR) and barcodes. Remark Office OMR™ included a quick statistics package, which we used to run initial reports even in real-time as the surveys arrived.

DISCUSSION

Topmost lessons had to do with survey design and cultural considerations. Clustered sampling necessitates trade-offs between cost and precision, for example, in the choice of size of clusters and the number of clusters sampled within strata. Translating and back-translating the questionnaire required us to pay attention to conversational and information-giving norms, including cognitive tasks such as judgment formation and response bias. Minimizing total survey error is made more vexing through quantifiable aspects, such as mode effects and coverage error, but also qualitative factors such as ethical considerations, translation, and cultural appropriateness, and even epistemology of research.

The detailed lessons learnt include the following:

- **Survey design.** There are trade-offs between cost and precision, for example, in the choice of size of clusters and the number of clusters sampled within strata.

- **Auspices.** Another often overlooked aspect is the patronage or institutional support in-country that guides the survey. Additional important features include religious sensibilities, political interests, and cultural and village authorities. The customary hierarchy of local authority becomes part of the project framework.

- **Budget.** Within the context of sometimes shifting sands, there can be a big gap between planned and implemented budget. A budget prepared well in advance must still take into account potential emergencies and unexpected economic and climatic calamities.

- **Work plan.** An operational work plan detailed down to the finest task is essential in monitoring or preventing possible sources of survey error.
• **Staff.** These surveys require years of experience in the project staff and these data cannot be collected by inexperienced researcher. Experienced enumerators are critical to the success of the survey.

• **Ethics.** Complying with the rules and regulations of the ethics procedures of the host institution is only one side of the question. Aligning these requirements with in-country ethical protocols is also essential.

**WAY FORWARD**

Given our recent experience of the cost and weight of paper surveys in humid climates, a switch over to PDAs or tablet PCs, using GPS, must be considered. There are some interesting possible advantages in switching to tablet PCs. First and foremost would be implementing a rapid and statistically valid survey so quickly that it could actually be used to measure entrepreneurial activity after a disaster or civil war—something that might be impossible with a paper or telephone survey. Voice and video recording capabilities can record interviews along with the data files. With global positioning system (GPS) capabilities, it would be possible to automatically record the time and place of the interviews, and even to track the progress of interviewers in real time.

Ideally, the next step would be a mixed mode experiment comparing GEM data collection by PDAs, paper questionnaires and telephone survey in a ‘borderline teledensity’ country. For control purposes, this experiment should be done in a geographically unified and ethnically uncomplicated developing country with teledensity in the 30% band. Candidate countries where no GEM surveys have been completed that satisfy these criteria are Vietnam with a 2008 teledensity rate of 32.6% and Costa Rica at 31.7% teledensity (International Telecommunication Union, 2010). Sample sizes of about 200 respondents under each treatment could be contemplated.

The comparative measures would include ability to collect, enter, and analyse data and calculate the impact of various forms of survey error. A sound experimental design would choose households without replacing them in the sampling frame. There would need to be more than one interviewer per method - preferably (for design efficiency) all interviewers would collect data using all three
methods. Interviewers would undergo similar training up to the point of data collection. The experiment would use data collection logs to document the proceedings of data collection and for use in reviewing the data. Interviewers themselves would then be interviewed to obtain information about their experiences. The experiment would be supervised by a technical consultant and an evaluation expert.

References


Ashmore, M. C., Pritz, S. G. & National Center for Research in Vocational Education (U.S.) (Eds.) (1983) PACE (revised) instructor guide, Columbus, Ohio, National Center for Research in Vocational Education.


