UNIVERSITEIT
STELLENBOSCH
UNIVERSITY

APPENDIX F

DATA MANAGEMENT PROCESS

Dr Chris van wyk

(University of Stellenbosch)

APPENDIX TO THE GRADE 3 IMPROVEMENT PROJECT OF THE UNIVERSITY OF STELLENBOSCH, SUPPORTED BY THE WESTERN CAPE EDUCATION DEPARTMENT AND THE WESTERN CAPE PROVINCIAL TREASURY

August 2010

Version: 10 August 2010

CONTENTS

Introduction.....1
Data management process 1
The Design of the Survey Questionnaires 1
Data collection instruments included classroom and school level instruments:.....2
Administering the instruments 3
The development of the capturing tool 3
Receipt of the completed instruments 4
Data Capturing and Cleaning.....4
Data Storage and Data Presentation.....5
Data Processing 5
DATA ANALYSIS 5

Educational policy improvements need evidence on what works and what does not, and (d)eveloping such evidence means regular high quality information about student outcomes must be generated

Hanushek: 2003

`…. But evaluating the appropriateness of alternative interventions requires proper investigation of the data, and the availability of such data.'

Van der Berg: 2009

INTRODUCTION

This Appendix describes the data management process of the school based survey conducted as part of the Grade 3 Improvement Project. The research team had to ensure the highest level of data integrity. The author of this Appendix was specifically tasked with leading this process.

The room for securing extra resources for basic education is limited. Education departments therefore have to ensure that current resources are utilized to the maximum. This requires good planning. Proper analysis of education data will go a long way in ensuring this.

The availability of applicable data and the variability of the data quality remain some of the biggest data challenges for evidence based policy making and sound data analysis. It was therefore of the utmost importance to ensure that data management process for the Grade 3 Improvement project contributed to the quality of the available data.

DATA MANAGEMENT PROCESS

Data management includes the range of activities to collect, capture, clean, share and integrate data from a number of sources. The data management process was based on the following clearly defined steps:

The Design of the Survey Questionnaires

Six different survey instruments were designed for use during the field work. The conception of the data collection instruments was informed largely by previous research and by the literature study conducted by Dr Ursula Hoadley. All instruments were designed to enable the maximum amount of quantitative data to be captured but at the same time allow important qualitative data to be gathered as well.

Data collection instruments included classroom and school level instruments:

Classroom level protocols for grades 1-3:

- Classroom observation schedule;
- Classroom document review instrument;
- Teacher interview schedule; and
- Grade 3 Learner multiplication task.

School level instruments:

- Principal questionnaire; and
- School library review instrument.

All survey instruments (except instrument 3, the learner test) consisted of a combination of three question types, namely

- questions that had different options (see Figure 1),
- questions that required narrative (qualitative) answers (See Figure 2), and
- questions that had a table format (See Figure 3).

Besides the development of the content, other important aspects, such as technical requirements and questionnaire layout had to be taken into consideration in the design of the survey instruments.

These formats also influenced the technical design of the database, therefore where possible the questions with the same format were grouped together and followed the same sequence on the survey instrument.

TABLE 1: **EXAMPLE OF QUESTION WITH OPTIONS**

How much grade-appropriate learning material (posters, maps and other visual aids) displayed on the walls **are learners able to see and read from their tables/desks**?		0=None 1=Fewer than 5 items 2=10 to 5 items 3=11 to 20 items 4=More than 20 items

TABLE 2: **EXAMPLE OF QUESTION WITH NARRATIVE ANSWERS**

If you have described discipline as poor in the question above, provide at least two examples of the kind of misbehaviour and/ or the cause.	
	Write description 98=Not applicable – discipline is not poor

TABLE 3: **EXAMPLE OF QUESTION WITH A TABLE FORMAT**

If there is any **delay in starting teaching a) at the beginning of the teaching day; b) after first break; or c) after lunch break**, describe what happens when timetabled teaching should be taking place. Estimate how much teaching or learning time in minutes it takes.		
1. *Describe delay (e.g. teacher late; teacher in a meeting; school feeding scheme delays teaching, etc.)* 0=None as teaching starts on time	2. *Describe what learners do (e.g. learners are idle i.e. no work done)*	3. *Write estimated time lost in minutes* 0=None as teaching starts on time 1=1 or less than a minute 99=Don't know e.g. fieldworker not present from the beginning of the session
a) Start of the day:		
b) After first break:		
c) After lunch/second break:		

Administering the instruments

The data gathering process is also dealt with in the report by Dr Cheryl Reeves (Appendix A). Of particular importance during this phase was ensuring that the fieldworkers were familiar with the instruments, including understanding the purpose of each instrument as well as the methodology for collecting the data using the specific instrument.

The training included collecting data in classrooms and thereafter reviewing the process as well as the quality of the collected data. This will be of particular importance where the WCED may wish to adapt and use the instruments in future.

The development of the capturing tool

In order to ensure that data were captured accurately and also speedily, given the number of questionnaires and the need for speedy feedback, the author designed a custom-coded

capturing tool. The technical layout and format of the questions on the survey forms played a key role in the development of the capturing tool. The development of the capturing tool had two distinct phases. The first part was the development of the database. The different tables in the database were derived from the types of questions on the survey instruments. The second part of the capturing tools entailed the development of the different input screens. The input screens had to have the same layout and sequence as the questionnaire. It ensured effective data capturing and speeding up the data capturing process. It further improved the capturing process and contributed to the process of ensuring that data was captured correctly and reliably.

Receipt of the completed instruments

All completed instruments received were first scrutinized to ensure that a complete set of data was received. All the forms received from the field workers were each assigned a unique identifier. The unique identifier linked the survey questionnaire to a school and grade. This link was important when the data were analysed. The unique identifier also made it possible to integrate the six survey instruments into one singled dataset and so as not to have to manage each dataset as an island of data.

Great care was taken to ensure that data sets were as complete as possible. As part of the ongoing feedback process conducted at the end of each week, potential problems in the completion of the instruments were identified and discussed with the fieldworkers.

All instruments were stored in separate files to enable easy retrieval for data capturing and record purposes. These will be kept by the research team at the University of Stellenbosch. The ethics of fieldwork demanded that fieldworkers had to sign an undertaking that they would not divulge any information obtained from individual teachers and principals to their joint employer, the WCED.

Data Capturing and Cleaning

The data was captured using experienced data capturers from the University of Stellenbosch. Great care was taken to ensure that the collected data from the instruments were captured accurately into the database. A system of cross-checking was developed by the author and the data capturers.

Data cleaning was another key step in the data management process to ensure that quality data was produced. Previous experience showed that unless great care was taken to ensure the quality of the data, the effectiveness of the data analysis could be greatly impacted by the lack of good quality data.

Data *verification* and *validation* were important processes during this step to prevent errors occurring in data. Verification and validation of data contributed to the quality of data.

Verification was used to prevent errors occurring when data was captured. Verification of data entailed checking that the data entered from the form correspond with the data in the database (proof reading). This process was managed with the assistance of students from the University of Stellenbosch.

Validation was used to check that the data entered has meaning, for example that data fell within a range of values.

Data Storage and Data Presentation

A database was developed to store the captured data in a format that enabled easy retrieval of the data. It also made data accessible for users in a systematic way. The data storage device also enhanced data integration with other datasets. Information is of far greater value when it is integrated. Data integration generally means linking different data sources through the use of a common field across a collection of data sources.

An electronic version of the database forms part of the final product will be available for further data analysis, as long as this does not conflict with the anonymity of respondents.

Key steps in the presentation of data:

- Integrate the data of the different instruments into one dataset.
- Format the data from relation database format into a file that can be imported into a spread sheet and statistical package
- Use the questions on the survey instruments as the label descriptions for the variables in the dataset

Data Processing

Questions with narrative type answers were processed in the following way:

- Capturing of the narratives
- Coding of these narratives per question per instrument
- Capturing of the codes
- Link codes to school and grades and integrate with other datasets

Drs Chris van Wyk and Cheryl Reeves were mainly responsible for this process. Given the extent of data (more than 800 variables) this was a mammoth task. However the availability of this data greatly enhanced the results and also presents opportunities for further research.

DATA ANALYSIS

The data analysis was carried out mainly by Prof van der Berg and some research associates as well as Drs Cheryl Reeves and Chris van Wyk.