

CHAPTER 3

PHASE 2: EFFECTS OF CURRICULUM CHANGE ON MEDICAL GRADUATES' PREPAREDNESS FOR INTERNSHIP: METHODOLOGY

3.1 THE RESEARCH FRAMEWORK

The reasons for the selection of a mixed method approach, using both quantitative and qualitative methods to seek answers to complex educational questions, was described in detail in the literature review in Chapter 1 under the discussion on complexity theory and the debate around sacrificing rigour for relevance. Although every effort was made in the research design to identify and take into account the confounding variables, such a study can never be fully controlled (Norman, 2002). The purely positivistic methods of natural science such as experimental design, randomised controlled trials and blinded trials were not possible, and would not have been helpful, in the study of the human behaviours involved in learning and practising clinical medicine. Cryer (1996) offers an interpretivist research paradigm as an alternative with an emphasis on exploration and insight rather than experiment. She points out that different parts of a complex investigation may need to be set in different research paradigms. Models help to simplify reality but if used alone they may lose intuitive insights but the use of metaphor can help to retain this information (Kernick, 2006). Thus a mixed method approach was chosen which used quantitative methods and statistical analysis where appropriate but which recognised the need for further qualitative exploration of underlying meanings and interpretations.

3.2 RESEARCH QUESTIONS FOR PHASE 2

The overall aim and four main research objectives for the study were given in Chapter 1. The first two objectives were dealt with in Chapter 2. The remainder of the thesis concerns the third and fourth objectives, namely:

Objective 3 to compare graduates of the traditional MBBCh curriculum and the GEMP curriculum against the model of intern competence using these instruments.

Sub-objective 3.1: comparison of overall questionnaire scores

Sub-objective 3.2: comparison in terms of hospital level allocation

Sub-objective 3.3: comparison in terms of population group

Sub-objective 3.4: comparison in terms of categories and items

The hypothesis (H_a): The graduates from the GEMP are better prepared for internship than the interns from the traditional medical curriculum

The null hypothesis (H_o): There is no difference in the preparedness for internship between the GEMP graduates and those from the traditional medical curriculum.

Objective 4 to relate intern performance to the content and methods of the relevant MBChB curricula experienced.

3.3 RESEARCH DESIGN

- A survey approach was used, conducted in the first four months of internship.
- It was a comparative, descriptive, cross-sectional study of two groups of medical interns from a single university, each group having graduated from a different curriculum.
- Collection of data was based on known (desired) outcomes in the form of a “Model of the Competent South African Intern” against which the two samples of interns were measured.
- Mixed methods research was employed as appropriate for the “exploration of complex initiatives and interactions among multiple players ... in natural settings” (Schifferdecker and Reed, 2009, p. 638). Both quantitative and qualitative methods were used to collect and analyse data.
- Results were examined from the perspectives of the various stakeholders (multiple voices) and were not combined. The reason for this decision is given later in this chapter.

It is important to remember that the outcomes measured in this study are the outcomes of the interns’ clinical encounters and not the patients’ outcomes such as improvement in

their condition or illness (Barrows and Tamblyn, 1980). Oppenheim (1966, p. 85) suggests that an up-front frame of reference, that tells the respondents the purpose of the study, will prevent them from “trying to guess the purpose of the questionnaire, and probably guess it wrongly”. This could affect the scoring quite considerably. This was an important aspect of the study and was appreciated by the respondents to the extent that every intern drawn in the sample agreed to take part, as did their supervisors, colleagues and patients.

3.4 METHODS AND INSTRUMENTS

Table 3.1 shows the methods considered for data collection in this study. The shaded columns to the right were eliminated as not feasible. Despite their obvious value as an object assessment of skills, OSCEs were not possible given the wide distribution of the subjects. HPCSA evaluation forms had only recently been introduced and the system was not fully operational during the first year of the study. The examination of intern clinical notes would have been an important objective measure and assessment forms were developed for this, but in practice it was impossible to find a supervisor with time to go through the notes and complete the assessment form during the researcher’s visit. Also several different doctors added to the notes and it was difficult to distinguish which had been written by the intern in question. Direct observation was adjudged not feasible given the time and travel limitations while an examination would have required the cooperation of several other medical staff in setting, marking and standard setting the examination. Journaling was unlikely to succeed as it required much time and effort from busy new interns. The following instruments were thus used in the study.

3.4.1 Questionnaires

The term questionnaire is used throughout this thesis for ease of reading although the format used was that of an inventory with a lead-in statement followed by 57 psychometric items which the respondents were asked to endorse using a 5-point Likert-type scale. A clear descriptor for each point of the scale was provided. A separate information sheet and consent form (Appendix B: B1) was designed for interns, supervisors, colleagues and patients.

Table 3.1 Operationalisation of complexity theory: Possible methods to assess intern competence in a 360° study - multiple methods are best (Holmboe and Hawkins, 1988)

	Self report questionnaire to interns on confidence & competence SELF REPORT RATING FORM Used	Interview with interns SEMI-STRUCTURED Used	Questionnaire to supervisors on intern competence RATING FORM Used	Interview with intern supervisor SEMI-STRUCTURED Used	Questionnaire to health team colleague - either nurse or intern peer RATING FORM AND COMMENTS Used	Interview with patient SEMI-STRUCTURED Used	Examination of Products of Work PATIENT RECORDS Not feasible, many interns contribute to notes. Unable to distinguish	OSCE/CEX Clinical evaluation, marked history & presentation STANDARDISED PATIENT Not feasible, interns too scattered	HPCSA evaluation forms Not feasible, forms not fully in use in 2006	Direct observation of interns Not feasible, interns widely dispersed in SA	Intern written journal Not feasible, dependent on intern cooperation	Written Exam Not feasible, exam security issues, need for staff input
	A	B	C	D	E	F	G	H	I	J	K	L
1. Theoretical knowledge Basic science, Pathology, Therapeutics, Pharmacology, Micro and Infection control	Y	Y	Y	Y								Y
2. Medical Problem solving and Clinical Judgement/Clinical competence	Y	Y	Y	Y			Y		Y			Y
3. Holistic patient management	Y	Y	Y	Y	Y		Y	Y	Y	Y	Y	Y
4. Community health	Y	Y	Y			Y	Y					
5. Professional Attitudes and Values/Ethics	Y	Y	Y	Y	Y	Y				Y	Y	
6. Effective Communication skills	Y		Y	Y	Y	Y	Y	Y		Y		
7. Working with others in a team	Y	Y	Y	Y	Y		Y			Y		
8. Self directed learning	Y	Y	Y	Y								
9. Confidence and personal attributes (intangible personal resources)	Y	Y	Y	Y	Y	Y				Y	Y	

The following questionnaires were developed from the Model which formed Phase 1 of the study and was described fully in Chapter 2.

1. A self-rating assessment for Wits medical graduates about their perceived preparedness for internship (Appendix B: B2). The frame of reference was for each intern to think back to how he or she felt at the commencement of internship, after the first few weeks of orientation (Objective 3). This was accompanied by an interview schedule to explore responses at the extremes of the scale and relate these to the type of curriculum followed (Objective 4).
2. The intern supervisors' rating of observed intern performance (Appendix B: B3). The frame of reference for supervisors completing the questionnaire for a particular intern was a comparison of the sampled intern with other interns that they had supervised over the last few years (Objective 3). This was accompanied by an interview schedule relating extreme responses to the type of curriculum (Objective 4).
3. The colleague rating of observed intern performance (Appendix B: B4). This questionnaire covered those items included in the previous two questionnaires which could be answered by a member of the health care team working in close daily contact with the interns. The two team members in closest contact with the interns were the nursing staff and fellow interns. The questionnaire was completed with whichever of these two categories was most appropriate in each setting. In certain institutions or departments the nursing staff did not work closely with the interns or were unavailable at the time of interview and here fellow interns were interviewed (Objective 3). Space was allowed for a brief written explanation of any of the responses if required. Some of these related to Objective 4.
4. The questionnaire to patients (Appendix B: B5) did not take the form of a rating scale to be completed by the patients themselves but was rather a semi-structured interview schedule based upon those items in the questionnaire that could best be answered by the patient (Objective 3). This method was selected to maintain uniformity of process

because several of the patients in the training hospitals were not conversant with questionnaires and some were illiterate. All of the interviews were administered by the same researcher. Open ended questions were asked and the answers recorded in the form of notes. An overall rating was given by the interviewer based on a set of subjective internal criteria (enthusiasm, the type of words used, body language and the level of detail offered). This was possible as the same interviewer conducted all patient interviews in both years of the study but these scores were not sufficiently robust to warrant the use statistical analysis.

3.4.1.1 Development and layout of the questionnaires

The Model of Intern Competence was developed from multiple sources and was validated by a panel of twenty seven experts as described in Chapter 2. Nine categories were identified and specific competency elements within each category were teased out and written as questionnaire items. There was no weighting of categories, but the number of specific items in each category was determined by the relative importance of that category as determined by the number of occurrence in the various sources used to develop the model.

The layout of the questionnaires was a folded A3 page with the information sheet printed on the first page. This gave the questionnaire a book-like feel with a logical sequence and print large enough to be easily read. No identifying information appeared on the questionnaires and a research code number was allocated to each intern with a different prefix for the intern (IRC), supervisor (SRC), colleague (CRC or PRC) and patient (PtRC).

Clear instructions were given for completing the questionnaire as well as descriptors for each of the five possible responses. The items were presented in a table format with a 5-point Likert-type rating scale requiring only a tick. The nine categories were not identified in the questionnaire. Interns were not given a column for “Don’t know” as they could respond regarding their known abilities and confidence to perform procedures. Supervisors and colleagues, however, were given this option as their responses depended upon their having

experienced a particular behaviour or having observed a skill. Signed consent forms were included for each respondent.

The interns' questionnaire comprised fifty seven items designed to elicit their perceptions of their own preparedness to function competently in each of the areas covered by the questionnaire. There were two open ended questions. The first of these asked whether there were any aspects of the curriculum, which the interns had learned and would have liked to practise but which were stifled in this due to the circumstances of internship. The nature of the internship year, with high patient loads and severe time constraints, may have prevented interns from practising some of the areas learned in their respective undergraduate curricula. The second open ended question asked for positive and negative reflections on the curriculum with an explanation of why this was so.

The supervisors' questionnaire used the same fifty seven items as that of the interns. The questionnaire looked the same as the intern questionnaire but the information sheet, introductory comments, descriptors for the ratings and items were phrased from a supervisors' point of view. The questionnaire explored the interns' preparedness, through the supervisors' observations of competence and performance of skills. The additional questions differed from those for the interns. Supervisors were asked to compare this particular intern with other Wits interns whom they had supervised in past five years and then to compare the intern with interns from other medical schools. A three point scale was used – weaker, similar or better.

The colleagues' questionnaire was a modified version of the supervisors' questionnaire and included only those items appropriate for nursing colleagues or peer interns to assess. The open ended item here merely asked whether there appeared to be any gaps in knowledge which could be attributed to the intern's undergraduate education. The patients' questionnaire was administered face to face and included only items relating to communication and interpersonal skills.

All questionnaires were scrutinised by two experts in the field of medical education for clarity, accuracy, readability and intuitiveness of the layout. The interview schedules were simple and were devised and checked by the two researchers who would use them.

3.4.1.2 Pilot testing of questionnaires

The intern questionnaires were administered to a pilot sample (convenience sample) of thirteen interns in 2005. They were asked to complete the questionnaire and to comment on the readability, any ambiguities which they had identified and their satisfaction with the information given and confidentiality of the survey. All respondents declared themselves satisfied with the information and confidentiality declaration, found the instructions understandable and the questionnaire clearly laid out and easy to complete. With only two exceptions, the questions were understandable and relevant and it took these interns between 15 and 25 minutes to complete the questionnaires. These responses contributed to the face validity of the intern questionnaires. The study design also allowed for clarification of meaning should any items not be clear to the respondents as the interviewer was present while the questionnaire was being completed.

Seven of the respondents to the pilot study wrote comments. Two requested that the wording of one of the items be made clearer as they had found it difficult to allocate a rating. This was the item on coping with uncertainty. One respondent suggested that an example be included in the item on minor surgical procedures – this was done and the words (eg. biopsy) were added in parenthesis. Another respondent suggested that, in the item on assisting competently with major surgery, the words “in the operating theatre” be added and the word “major” be removed. These minor changes were made as they did not effect the meaning of the items and increased the clarity. Three (3) respondents commented on how difficult it was to manage the many patients for whom they were responsible during internship. One wrote that the item on coping with the long hours needed to be expanded to include coping with the patient load.

“Not just coping with long hours but need to have a more definite item on the workload. Last year we thought we were dying looking after 2 patients. Now we have 20 and it’s really terrible. The workload is very stressful”.

This suggestion was accommodated by adding a separate item on workload as it had a different meaning from long hours and seemed to be important to the interns themselves.

An important point regarding the timing of the research emerged:

“It is clear that the questionnaire refers to when I first started at the JHB hospital. The first month was hectic. I remember it well. On call first day in trauma and I didn’t know the equipment or how to connect it up. Patients were critically ill. It was terrible! Would have answered this questionnaire a bit differently if it had been handed out in January or February”.

It was decided, on the basis of this comment, that the questionnaires and interviews should take place only once interns had received orientation to their new environment and had settled into their rotations. They could then be asked to think back to their early weeks.

The supervisors’ questionnaire was given to 5 pairs of supervisors in 2005 as part of a pilot study and test for inter-rater reliability, discussed in section 3.4.1.4. Each pair was asked to select an intern that they had both supervised and to complete the questionnaire independently for that intern. They were asked to comment on any difficulties that they had experienced in using the forms. All pairs agreed that the questionnaire was relevant and the wording clear and this helped to establish construct validity of the questionnaire items. They had no difficulty completing it but noted that they could only respond to items that they had actually experienced or seen themselves resulting in many missing items.

The colleague and patient questionnaires were not piloted as both were administered by the researcher who read out the items and recorded the responses. Any queries could be explained at the time of completion. The interview schedules were also not piloted as they were used

only by the two interviewers involved in the study who could clarify any misunderstandings at the interview. There were no problems encountered with the use of these schedules.

3.4.1.3 Validity (Quantitative)

1. Face validity of the questionnaire items was achieved by having two educational experts read the questionnaire, followed by validation by interns during the pilot study in the year prior to the commencement of the study. These respondents had little or no trouble completing the questionnaire. Content validity was achieved by the rigorous process of examining local and international literature and interviewing a sampled group of experts during the development of the Model of the Competent South African Intern. This served to select those areas which were considered important to competence for all South African interns. The model was then validated by an independent panel of medical and educational experts.
2. Concurrent validity was achieved through triangulation which is defined as “the use of two or more methods of data collection in the study of some aspect of human behaviour” (Cohen, Manion and Morrison, 2007, p. 141) which enables the cross-checking and corroboration of results. Many other proponents of naturalistic inquiry also recommend the use of triangulation (Delamont, 1992; Denzin, 1978; Hammersley, 1990 and Fetterman, 1989). In order to validate this study methodological triangulation (Cohen *et al*, 2007, p141) was used to obtain data on intern preparedness by employing three different techniques, psychometric responses to items on a Likert-type scale, open-ended written comments and semi-structured interviews.
3. Construct validity “Every hypothesis tested in research is constructed in a theoretical context of ideas” (Graziano & Raulin 1993, p.171) Construct validity in this study would depend upon how well the results support the theoretical framework upon which the research was based and whether this was indeed the best theoretical explanation of the results.

“To establish construct validity the researcher would need to be assured that his or her construction of a particular issue agreed with other constructions of the same underlying issue ...” (Cohen *et al*, 2007, p. 138). In this study, competence and complexity are understood in the light of the literature on the subject and have been clearly defined. Validation of the model upon which the research instruments were based demonstrated “that the categories that the researchers are using are meaningful to the participants themselves” (*ibid*, p. 138). The pilot study ensured intern understanding.

(Schifferdecker & Reed, 2009, p. 638) suggest that mixed methods research, which employs both quantitative and qualitative methods can “inform the development of one method from another, using the methods sequentially for the purposes of increasing construct validity”. This was done in the intern study through the use of a questionnaire (quantitative) followed sequentially by an interview (qualitative).

4. External validity, in the positivist sense of generalisability to the wider population, is permissible only to other Wits interns in each of the study years respectively. However, in naturalistic enquiry Cohen *et al*, (2007) refer to various authors who interpret generalisability as “comparability and transferability” in the sense that sufficiently in-depth descriptions enable others to decide how far the findings from the research are generalisable to another situation (*ibid*, p. 137). Such an interpretation may allow some of the results from this study to be generalisable to interns who have graduated from similar curricula in other South African universities.

3.4.1.4 Reliability (Quantitative)

It was not possible to test the reliability of the questionnaire using a test-retest technique due to the wide geographical distribution of the intern respondents and the “practice effect” (Hubbard, 1978, p. 60). An alternative approach, which required only one administration, was the split-half test of reliability. This test was performed by scoring the odd and the even numbered items separately in the 2006 intern questionnaire and looking at the correlation

between the halves. The questionnaire was sufficiently long (fifty seven items) to allow for the use of this technique. Streiner and Norman (1995) advise that when looking at a trait or behaviour, in this case “competence”, all of the items should relate to that attribute, known as homogeneity. There should be moderate correlation between items in a scale. If the correlation is too high there might be redundancy. The Cronbach’s alpha correlation coefficient between the two halves was calculated ($\alpha = .95$). This score is somewhat high although redundancy might have been introduced in the questionnaire by breaking the question on clinical procedural skills down into individual skills. However, there appears to be sufficient reliability to have faith in the instrument. Cronbach’s alpha was the appropriate test for this questionnaire because there were more than two response alternatives.

Consistency of measurement (reliability) is also gained through attempts to control the many confounding variables that might cause differences between the two groups of interns other than that brought about by the change in curriculum. These controls are described in the section on matching the samples. It is also strengthened by the breadth of the questionnaire, based upon a carefully researched model, which serves to reduce the sampling error (Barrows and Tamblyn, 1980).

Reliability as equivalence (Cohen *et al*, 2007, p 146) was demonstrated by a small inter-rater reliability test conducted on the supervisor questionnaire during the pilot phase of the study described under section 3.4.1.2. Synchronic reliability is achieved as far as possible by “seek(ing) similarity of data gathered in the same time” (Cohen *et al*, 2007 p. 142). This aspect emerged as important during the pilot testing and was one of the criteria for selecting the matched sample in 2007.

The supervisors’ questionnaire was given to 5 pairs of supervisors as part of the pilot study and each pair was asked to select an intern that they both knew and to complete the questionnaire independently for that intern. Data from this pilot study were used to test the inter-rater reliability. Items where at least one of the supervisors in a pair had not given a response were counted as “missing” items. Sixty one (61) items were classified as missing leaving 219 items with paired responses. Of these, 191 items were given the either the same

score or a score within one point by the pair of supervisors responding to the item. Using a simple formula in Cohen *et al*, (2007, p. 147) the inter-rater reliability was calculated.

$$\frac{\text{number of actual agreements}}{\text{number of possible agreements}} \times 100 = \frac{191}{219} \times 100 = 87\%$$

This level of consensus (87%) was sufficient to accept the ‘equivalence reliability’ (Cohen *et al*, 2007, p. 147) of the supervisor questionnaire.

3.4.2 Interviews

A short half hour interview was conducted with each of the sampled interns and his or her supervisor directly after the questionnaire had been completed. The purpose of these interviews was to meet Objective 4 by relating the preparedness of each intern to the content and methods of his or her particular curriculum. Items where scores were given at the extremes of the scale (1 or 2 and 5) were followed up in the interviews to clarify the relationship of these scores to the interns’ MBChB curriculum.

The semi-structured interview schedules (Appendix B) for use with interns, their supervisors, colleagues and patients were developed to ensure that the process and instructions for each interview were identical for every respondent and with opportunity for open ended comment. Responses were written up by hand during the interview in the form of notes and typed up as soon after the interview as possible so as not to lose any of the detail. A neutral and very experienced interviewer (who was not the researcher) conducted the interns’ and supervisors’ interviews in both years of the study. The reasons for this are discussed under the section on bias. Interviews with health team colleagues and patients were conducted by the researcher as these respondents were not known to her. If Afrikaans was the preferred language, the questions were translated by the researcher and responses recorded in the patient’s language.

A global score was allocated by the researcher, using the following rubric:

- Score = 1 or 2: the patient did not answer fully or seemed unsure of what to say
- Score = 3: the patient answered all questions but did not elaborate
- Score = 4: the patient answered all questions fully giving some examples
- Score = 5: the patient answered all questions fully, volunteered additional information to clarify and supported comments with anecdotal evidence.

3.4.2.1 Validity and trustworthiness in qualitative research

Lincoln and Guba (1985) and Pretty (1995) both suggest that in qualitative research the concept of trustworthiness of research findings is established by demonstrating, in qualitative terms, the four elements of trustworthiness. These are credibility, transferability, dependability and confirmability and they parallel internal validity, external validity, reliability and objectivity respectively. Credibility is determined largely by the researcher and is enhanced if the researcher is familiar with the phenomenon and setting being studied, and is able give a content-rich, plausible account through skills in drawing people out, using a multidisciplinary approach and triangulation (Miles and Huberman, 1994). Transferability is limited in qualitative research but findings can relate to other contexts. Dependability involves accommodating change as the study progresses and the researcher's understanding becomes more refined and finally, although qualitative research is inherently subjective, Miles and Huberman (1994, in Clifford and Serlin, 2006, p. 417) frame confirmability as "relative neutrality and reasonable freedom from unacknowledged researcher biases". All of these elements of trustworthiness have been addressed to a greater or lesser extent in this study as explained in the preceding paragraphs.

Cohen and Manion (1989, p. 318) attempt to align qualitative and quantitative validation by proposing that "one way of validating interview measures is to compare the interview measure with another measure that has already been shown to be valid. This kind of comparison is known as convergent validity. If the two measures agree it can be assumed that the validity of the interview is comparable with the proven validity of the other measure". The use of psychometric responses followed by interviews to tease out detail and relate the responses to curricula structure and content would fall into this category.

Pretty (1995, p. 1251) suggests that "it is clearly time to break the domination of the old paradigm of positivism for science, and so explore the alternatives".

3.5 STUDY POPULATIONS AND SAMPLING

The research populations comprised all Wits medical graduates who were registered to commence their Internship Training Programme in the two study years. The annual lists of intern allocations for all Wits graduates in 2006 and 2007 were obtained from the Department of Health in Pretoria.

The interns in 2006 were graduates of the final year of the traditional MBBCh curriculum who commenced their internship in January 2006 and were on a one year internship programme. All had passed the final MBBCh VI examinations and had attended Wits for at least the final four years of the degree. Medical students who had to repeat examinations and therefore only commenced their internship mid-year were excluded from the sample in 2006.

The interns in 2007 were the first cohort of graduates of the new Graduate Entry Medical Programme (GEMP) who commenced their 2 year internship programme in January 2007 or had repeated only a single six week rotation. This change was necessary in order to match the sample to that of 2006. All had passed the final MBBCh VI examinations and had attended Wits for at least the final four years of the degree

3.5.1 The study population in 2006

Figure 3.1 shows the breakdown of the 2006 Wits interns by population group. As explained in Price & Smuts (2002) it was important to identify the opinions of Indian, White and Black/Coloured students separately to ensure that the changes in the medical curriculum would not further disadvantage groups that had already been disadvantaged through the historical, racially based educational and socio-economic systems of apartheid in South Africa.

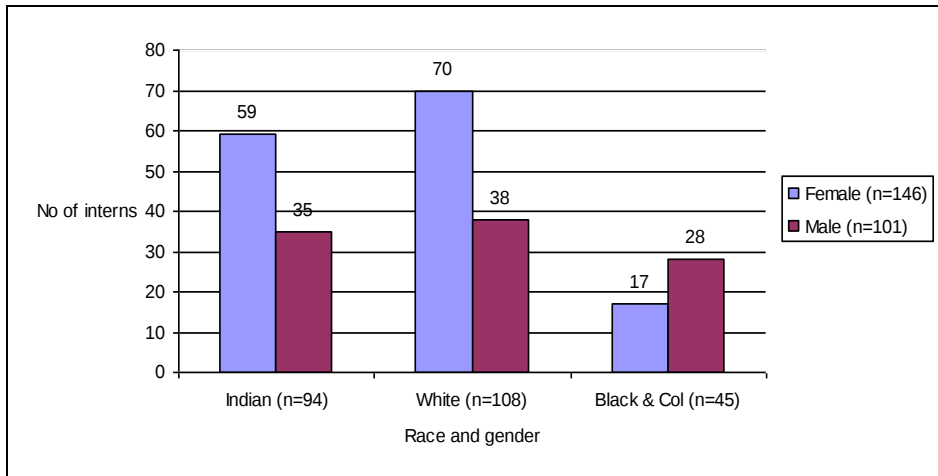


Figure 3.1 The Wits intern population in 2006 (n=247) by race and gender

Figure 3.2 gives the South African Department of Health's allocation of Wits interns to hospitals at various levels. Where allocations were to complexes which included a tertiary hospital with satellite institutions at other levels, the level of the institution where the sampled intern commenced his or her internship was recorded for purposes of data analysis.

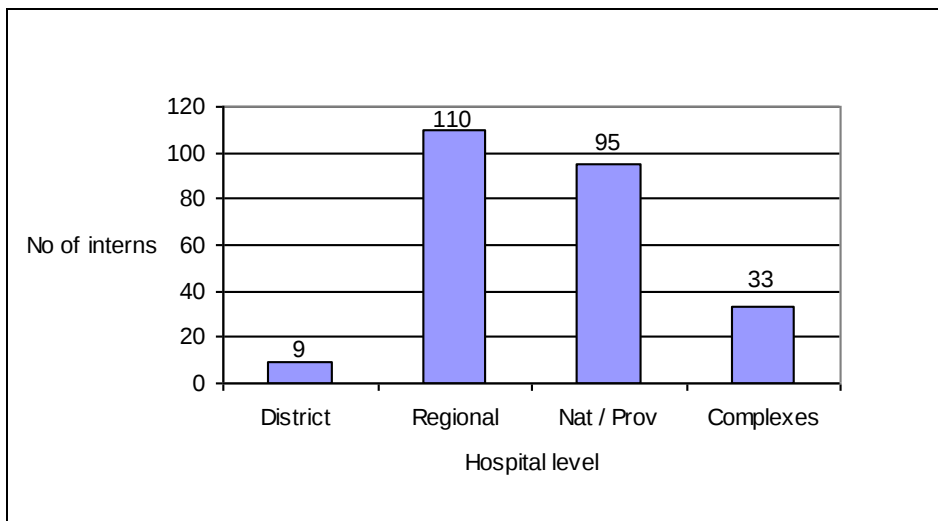


Figure 3.2 Wits Intern allocations to training institutions at the various hospital levels in 2006 (n=247)

Interns are allocated by the Department of Health to approved intern training facilities in the nine provinces of South Africa and also Namibia.

Figure 3.3 presents the distribution of the 2006 Wits intern population in the various provinces and the map in Figure 3.4 shows the geographical location of these provinces in South Africa and also of Namibia. Although interns were given a choice as to where they would like to do their internship they did not always get their first or second choices, some even having to accept their fifth or sixth choice. The interns drawn in the research sample were visited in whichever province they had been allocated.

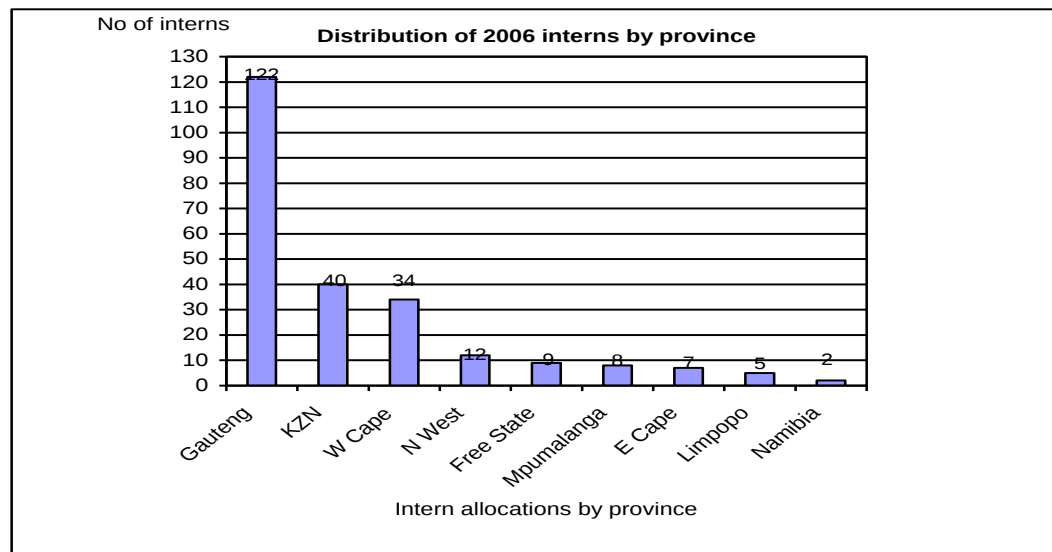


Figure 3.3 Distribution of the 2006 intern population by province (including Namibia)



Figure 3.4 Map of the nine South African provinces and Namibia
(www.places.co.za)

3.5.2 Sampling methods and sample sizes

3.5.2.1 Sampling the traditional curriculum graduates in 2006

Time, personnel and financial constraints precluded the follow-up of all the 2006 interns and the class size was too small for cluster sampling to produce a representative sample, although this would have saved travel time and expense. The probability sampling method selected was therefore stratified random sampling (Graziano & Raulin, 1993; Cohen *et al*, 2007, p.110).

Table 3.2 shows that there were few Wits interns allocated to district level hospitals, yet it was important to know whether the curriculum prepared interns to work competently at this level. Likewise there were fewer Black and Coloured graduates than the other population groupings and it was important that their voices be heard. A purely random sample might have omitted these data and so a stratified sampling frame was drawn up with race and hospital level as the strata and a proportional sample was drawn randomly within the strata (see Table 3.3 below).

It also allowed for oversampling of interns in district level hospitals and the Black and Coloured categories which were under-represented in the population.

The population size was 247 and the required minimum sample size was sixty one (61) for a 95% confidence level with an expected frequency for completing of 70% and a worst acceptable level of 60%. This was calculated using the Stata IC/10.0 Statistical Package. (Formula: Sample Size = $n/(1-(n/\text{population}))$: $n=Z*Z(P(1-P))/S*S$) (Kish, 1965).

Oversampling of District level hospitals (x3) and Black and Coloured interns (x2) brought the required sample size to seventy five (75). For practical purposes an additional 10% sample was drawn separately to be used in the case of non-response or sampled interns being unavailable. The Stata package provided random numbers within each cell and these were used to draw the sample. The final sample size for the study was seventy six (76) interns.

Table 3.2 The sampling frame for all Wits interns in 2006

	Black & Coloured	Indian	White	TOTAL
Level I - District	4	4	1	9
Level II - Regional	30	40	40	110
Level III - Nat/Prov	9	40	46	95
Level IV - Complexes	2	10	21	33
TOTAL	45	94	108	247

Table 3.3 The sample drawn from the 2006 intern sampling frame, including 10% “spare” respondents (Stata IC 10.0)

	Black & Coloured*	Indian	White	TOTAL
Level I – District**	3	3	1	7
Level II - Regional	17	11	11	39
Level III - Nat/Prov	4	11	12	27
Level IV - Complexes	1	2	6	9
TOTAL	25	27	30	82

* = (column) over-sampled x2

** = (row) over-sampled x 3

3.5.2.2 Sampling GEMP graduates in 2007

In 2007 it was essential to control for as many threats to internal validity as possible (Graziano and Raulin, 1993) so that comparisons of preparedness could be measured between graduates

of the 2006 and 2007 curricula. “The common method of constructing probabilistically equivalent groups is matched sampling” (Quigley, Muñoz and Jacknowitz, 2003 p. 5). The same sized sample (76 interns) was selected purposively from the 2007 intern group to match as closely as possible each respondent in the 2006 sample. Four important criteria were used for matching the samples thus attempting to eliminate these factors as a major source of confounding. These were:

- Hospital level (subsumes intern environment). Wherever possible the same hospital was used. If no exact match was achievable then an equivalent hospital at the same level and rural/urban position was selected.
- Race (subsumes to some extent educational advantage or disadvantage, socio-economic status and language)
- Timing of the survey visit – surveyed within four weeks of each other (subsumes experience and confidence gained)
- Quartile ranking in the MBBCCh VI final year mark (subsumes personal ability and application to undergraduate studies).

Additional variables such as gender and age could not be included for matching purposes as the sample sizes were not large enough to find exact or close matches using too many variables. The four chosen were considered the most important for the purposes of this study.

Table 3.4 shows that 76% of the intern pairs were matched on at least three of the four criteria, and the remaining 24% matched on two of the four criteria.

Table 3.4 Percentages of intern pairs in 2006 and 2007 which fitted the four criteria

Matching code	Frequency	Percentage	Cumulative %
4 out of 4	30	39.5	39.5
3 out of 4	28	36.8	76.3
2 out of 4	18	23.7	100.0
1 out of 4	0	00.0	
TOTAL	76	100.0	

3.5.2.3 Selection of supervisors, colleagues and patients

For each intern sampled, the supervisor most closely involved with that intern in the first weeks of the rotation was asked to participate. The interns were requested to nominate two or more health team members, either nursing staff or intern peers, and at least two patients for whom they were responsible and who were well enough to be interviewed. The researcher then selected one colleague and one patient to complete the questionnaire and/or interview, the final selection being determined by availability and circumstances at the time of the visit. The final selection of colleagues is given in Table 3.5 and illustrated in Figure 3.5. The significant difference (Yates $\chi^2 = 2184$, df = 1, **p<0.001**) between the numbers of intern peers and nurses in the two years is discussed further in Chapter 5.

Table 3.5 Comparison between the proportion of intern peers to nurse colleagues who completed the questionnaires and interviews in 2006 and 2007

2006			2007		
Intern peer (%)	Nurse (%)	Missing (%)	Intern peer (%)	Nurse (%)	Missing (%)
19 (25%)	53 (70%)	4 (5%)	45 (59%)	29 (38%)	2 (3%)
TOTAL = 76 (100%)			TOTAL = 76 (100%)		

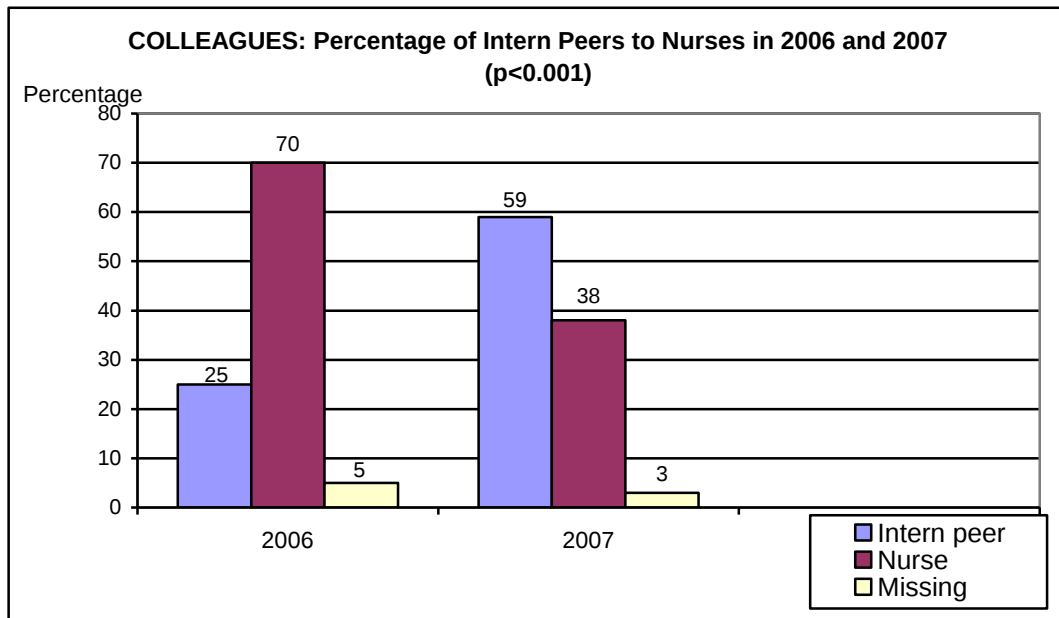


Fig 3.5: Proportion of intern colleagues to nurse colleagues in 2006 and 2007

3.6 DATA HANDLING

3.6.1 *Questionnaires and interviews*

Permission to conduct the research at the various hospitals was required. This entailed seeking written permission first from the head of the Department of Health for each of the provinces and then from the medical superintendent, chief executive officer or manager of every hospital visited. In three provinces the research proposal had to be approved by their own provincial research Ethics committees (Gauteng, the Western Cape and Mpumalanga). The process of obtaining these permissions took almost eighteen months to complete. Biographical data on the interns were available, with written permission, from the Wits student records.

The questionnaire administration and interviews with interns and their supervisors were conducted by a professor of Chemical Pathology who had been out of South Africa for five years, working on a new curriculum at a foreign university. She had therefore played no part in the introduction of the new GEMP curriculum and was not known to the interns. She was considered neutral and unbiased, yet knowledgeable, discreet and with sufficient standing in the medical profession to be accepted by these respondents.

Colleagues and patients were not known to the researcher and so their questionnaires and interviews were conducted by the researcher herself. Potential bias was controlled as far as possible by asking identical questions of each patient or colleague from the respective questionnaire or interview schedule and recording the responses *verbatim* in writing. No training of interviewers was necessary as each was experienced in this field and each collected data from the same groups of respondents in both of the study years.

The researcher and interviewer planned their trips to the various provinces as strategically as possible to save on costs. The sampled interns were contacted by telephone before traveling to the various hospitals to explain briefly the reason for the study and to request the interns' agreement to participate. A meeting was scheduled in the hospital in which they were

working, at a time and place convenient to both the intern and his or her supervisor. No interns refused to participate and even those who were a little apprehensive at first were reassured after reading the information sheet and having their questions answered by the interviewer.

A preliminary meeting was held, if possible with the intern and his or her supervisor together, to explain the study and obtain signed consent for each part of the study on a separate consent form. A self-report questionnaire on “preparedness for internship” was given to the intern with a request that it be completed immediately and handed back to the interviewer. A similar questionnaire on “observed competence of the intern” was given to the supervisor with a similar request that it be handed back to the interviewer on completion. The post-questionnaire interviews took about half an hour each and were conducted with the intern and supervisor separately to relate their questionnaire responses which fell at the extremes of the Likert scale back to the undergraduate curriculum.

While these interviews were taking place, the researcher met with the selected colleague and patient to administer these questionnaires and conduct the interviews. The intern introduced the researcher to these respondents personally so that there could be no confusion as to which intern they were discussing and to ensure that the colleague or patient understood that the research was being undertaken with the interns’ knowledge and consent. Each colleague and each patient also signed individual consent to participate.

Where a face to face meeting with an intern or supervisor was impossible due to the temporary absence of one or the other of the respondents at the time of the researcher’s visit, a telephonic interview was arranged. This occurred in approximately ten cases in each of the study years, mainly with supervisors. The information sheet explaining the study was read out exactly as it appeared on the front of the questionnaire form and verbal consent was given by the respondent and signed for by the interviewer on his or her behalf. The questionnaire items were then read out and the responses recorded on the questionnaire sheet. The post-questionnaire interview was completed over the telephone in exactly the same manner as in the face to face interviews.

3.6.2 *Data reduction (capture, cleaning and coding of data)*

Sets of questionnaires (intern, supervisor, colleague and patient) were filed together in plastic sleeves using the Intern Research Code (IRC) number and prefixes as described earlier in this chapter. Interview notes were typed up as soon after the interview as possible to ensure accuracy and transcripts were filed electronically and also as hard copy in the research files.

Biographical and response data were entered into a Microsoft Office 2003 Excel spreadsheet. A coding table was created and all data, including biographical data and scores, were represented numerically. Initial frequencies were calculated for each questionnaire item and these were used to check the data for any inaccuracies or missing data and corrections were made. The cleaned data file was then used for all analyses.

3.6.3 *The concept of a 360 degree study and the limitations encountered*

Leung (2002, p. 694) highlighted the distinction between clinical competence (what the doctor can do) and clinical performance (what doctors do do) and stated that neither is value free. In the current study the possibility existed that interns reported subjective feelings of confidence in their undergraduate preparation to perform procedures, make judgments and apply the knowledge they had gained. The biases here would include how self critical they were, their insight into their own performance, blaming the curriculum for shortfalls in knowledge or experience, whether they had taken advantage of experiences offered, as well as personality, upbringing and cultural differences.

The supervisors and colleagues, on the other hand, could report only what they had observed. This may have led to different biases such as the “halo effect” (carrying over the observation of “some” well or poorly performed procedures to all procedures) (Grey, Pearson, Rolfe, *et al*, 2001) or reporting on intern competence in general rather than that of the particular intern in question. Personal differences or prejudices, differing clinical protocols, personality issues and differences between the interns’ undergraduate education and their own medical training

are possible. Rolfe, Andren, Pearson, *et al* (1995) suggested that a supervisor from the same medical school might rate interns more positively than a supervisor trained at a different medical school. Patients could also only report on observed or experienced competence.

Figure 3.6 depicts the different perspectives on intern competence included in this study, together with the possible biases that could occur for each of the respondent groups. This serves as a reminder of some of the difficulties inherent in research on complex human behaviours, where technical rigour must sometimes be forsaken in order to fully explore areas of human relevance as suggested by Schön (1983) in his “swampy lowlands” (shown in Figure 1.3 in Chapter 1).

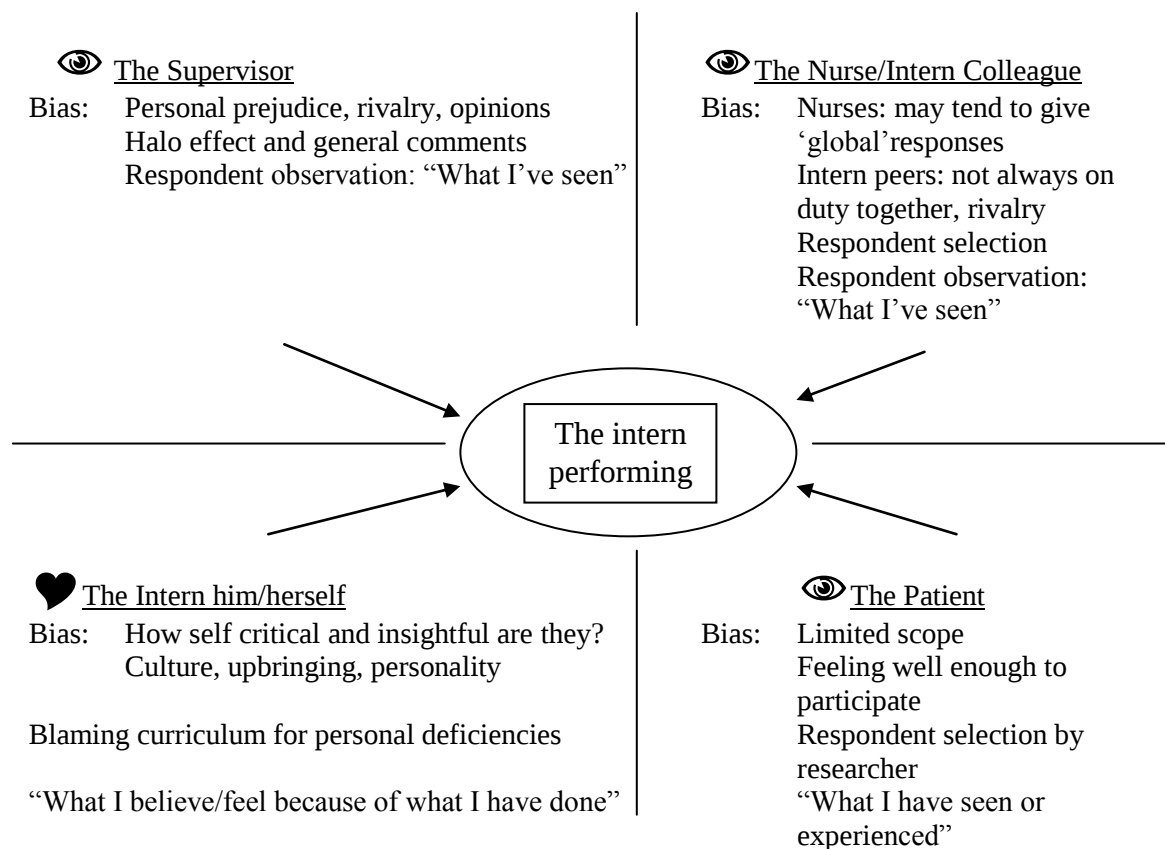


Figure 3.6 Diagrammatic representation of the four perspectives on intern preparedness measured in the 360⁰ study

In summary, Figure 3.6 highlights the fact that interns reported on how competent they felt while the other respondents reported on the competence that they saw or experienced. This led to a decision to treat each group of respondents separately and was supported by other studies, especially one by Kegel-Flom (1960) which also found that the performance of a medical intern was viewed differently by supervisors, peers and the intern him or herself. In fact, four perspectives are likely to enrich one's understanding of interns' competence more than just one.

3.7 DATA ANALYSIS

3.7.1 Statistical analysis of quantitative data

There has been much discussion in the literature about the legitimacy of treating ordinal data as if it were interval data and then using parametric statistical techniques to analyse the results (Carifio & Perla, 2008; Jamieson, 2004; Knapp, 1990; Kuzon, Urbanchek and McCabe, 1996). Since the data were ordinal it was decided that in the interests of accuracy non-parametric statistical techniques would ordinarily be used for the quantitative data throughout the study. All statistical analyses were conducted using PC SAS version 9.1 (SAS Institute Inc. Cary, NC, USA).

The interns had undergone six or more years of undergraduate medical training and, not surprisingly, it was found that very few responses fell into response category 1 (Did not know enough/not prepared or not confident/competent to perform) on the five point rating scale. It was therefore decided to combine this category with response category 2 (Rather unsure/seems hesitant or needed guidance through the process/competent under close supervision) for all final analyses, resulting in four response categories instead of five for the Intern and Supervisor questionnaires. There were so few low scores given by Colleagues (nurses and peer interns) that it was decided to combine categories 1, 2 and 3 resulting in only three response categories for these respondents.

Cochran-Mantel-Haenszel Statistics (MH Chi-Square) using Modified Redit Scores, a procedure based on the Wilcoxon-Mann-Whitney test which removes the difficulties in testing or forming confidence intervals and results in a more informative summary of the data (Flora, 1974), were used to test for significant differences between the 2006 and 2007 responses. This is a non-parametric test for measuring ordinal variables on two independent populations of students (i.e. not the same group before and after exposure to curriculum change) (Walker 2002, McDonald, 2008). Where the direction of the difference was difficult to determine, a further *post hoc* analysis was carried out by the consultant statistician using 2x2 tables, in addition to the standard 2x4 tables and applying the Yates correction and Bonferroni correction. However, these 2x2 tables did not always help in clarifying the general direction of the difference between the two years, only the differences between each of the four discrete response categories for items on the Likert-type scale. A 95% confidence interval (CI) and probability level of 0.05 were used throughout.

The Fisher's Exact Test was performed where expected counts of less than five were recorded for certain cells. This test is an alternative to the chi-square test for comparing two independent binomial proportions and is useful when the normal approximation to the binomial might not be applicable, such as in the case of small cell sizes. It compares each Likert-type response category in 2006 with that in 2007 but does not take ordinality into account (Walker, 2002; McDonald, 2008).

The format for presenting the comparative results was as follows:

- Overall scores for interns, supervisors and colleagues (where applicable) in 2006 and 2007
- Overall scores by hospital level and ethnic group
- Scores for interns, supervisors and colleagues for each of the nine categories of the Intern Competence Model (see Table 3.6)

Table 3.6 Individual items falling under each of the nine categories

Nine Categories of The Intern Model	Questionnaire item numbers included in each category
Category 1:	1,2,3, 4
Category 2:	5, 6, 7, 8, 9, 10, 11, 12, 13, 14
Category 3a: (*Category 3b:)	18, 28, 30, 31, 32, 33 (15, 16, 17, 19, 20, 21, 22, 23, 24, 25, 26, 27,29)
Category 4:	34, 35, 36, 37
Category 5:	38, 39, 40, 41
Category 6:	42, 43, 44, 45
Category 7:	46, 47, 48, 49
Category 8:	50, 51, 52, 53
Category 9:	54, 55, 56, 57

*Note: Category 3b consisted of scores for individual procedures (not included in overall Cat 3a score)

- The 57 items were then analysed individually for interns, supervisors and colleagues. Only the scores for items that showed significant differences between the traditional curriculum and GEMP graduates were presented, together with their qualitative data, in order to better understand the differences as they related to undergraduate preparation in the two different curricula. Patient data were dealt with qualitatively only.

3.7.2 *Thematic analysis of qualitative data*

Cohen *et al* (2007) describe three main paradigms from which to approach the study of behaviour. These are:

- the normative or positivist approach which uses the model of the sciences is impersonal, objective, conducted from the outside and tries to explain behaviour and seek causes, generalising from the specific.
- the interpretive approach which is non statistical and is concerned with the individual. It includes phenomenology, ethnography and research with a more practical interest. It is subjective, with personal involvement of the researcher and seeks to understand actions and meanings rather than causes, interpreting the specific rather than generalising from the specific.
- the critical approach which includes action research and critique of political and ideological factors and power that shape behaviour.

Although this study has taken a normative approach using statistical methods and has attempted a degree of objectivity, this is cannot be fully successful in the study of human behaviours as complex and intangible as those involve in the study of intern competence.

The semi-structured interviews with interns, supervisors, other health team members and patients served to better understand the quantitative scores given in the questionnaires and added an important further dimension to the study of intern competence. Denzin & Lincoln (2003, p. 4) describe qualitative research as “a situated activity that locates the observer in the world” and go on to explain that at this level qualitative research involves study in a natural setting. It is this interpretive, naturalistic approach to the world which attempts to make sense of phenomena in terms of the meanings that people bring to them. It was in this spirit that the interviews were conducted, a means of hearing the meanings behind the psychometric responses to the questionnaire items.

The qualitative data took the form of written responses to open ended questions for the interns and supervisors as well as transcripts of interviews held immediately after completion of the questionnaires and with patients. All field notes taken during the interviews were typed out in full as soon after the interview as possible to facilitate accuracy.

Qualitative data reduction is an integral part of analysis and this study followed the Interactive Model described by Miles and Huberman (Miles and Huberman, 1984). It was an ongoing process which involved different levels of decision-making in the reduction or focusing process. Pre-data collection reduction, also called “anticipatory reduction” (*ibid*, p. 24) was planned into the conceptual framework, the research questions, the sampling methods used, the design of the interview schedules and the decisions on how much qualitative data to collect and how “open” to leave the questions. The major reduction took place when analyzing the material vertically. Comments for each item were analysed text by text and the volume reduced by counting repetitions and identifying emerging themes (van Zyl & Bowman, 2007). This process is sometimes called progressive focusing in some qualitative research circles (Miles and Huberman, 1984, p. 4). The qualitative analysis served to better understand the

reasons for success or failure of aspects of each curriculum to prepare the graduates for internship (and thus to meet Objective 4). During qualitative analysis it was important to record not only the recurring themes but also interesting content that was noticeably different or presented novel ideas. These themes were quantified where appropriate and recorded in table form as part of the vertical analysis process.

The next step was a horizontal analysis to compare and discuss the similarities and differences within the themes which emerged from the 2006 and 2007 respondents. Conclusions were drawn and the findings were compared with the key issues in the literature on change in medical education (van Zyl & Bowman, 2007). It was important always to maintain the “voice” of the respondents and for this reason relevant quotations are included and are used as the principal form of data reporting. In certain instances, particularly in the analysis of the patients’ comments, some level of language-sensitive discourse analysis was employed. The resulting qualitative evidence was used to complement and qualify the statistical results and to add additional layers of understanding in meeting the research objectives, particularly Objective 4 which dealt with relating the research findings on intern competence and performance to the content and methods of the two different curricula as experienced (or perceived) by the 2006 and 2007 respondents.

During the post questionnaire interview, interns and supervisors were given the opportunity to add any further general comments. Pertinent themes arising from these comments are reported in the results section.

3.8 ETHICAL CONSIDERATIONS

Ethics clearance for the research was sought and obtained in the first instance from the Committee for Research into Human Subjects (Medical) of the University of the Witwatersrand (Clearance number: M050453) as well as the Faculty of Health Sciences Research Ethics Committee of the University of Cape Town (REC REF: 210/2005) and the Mpumalanga Department of Health Research and Ethics Committee (09/06/2005). Letters of permission to conduct the research in the various provinces were obtained from the offices of

the Directors General of Health, as well as written permission from the Chief Executive Officers, Medical Superintendents or Intern Curators from each of the hospitals visited. This entire process took more than a year to complete.

It was essential to identify the Wits MBBCh graduates in the sample in order to visit them at their internship facility and to analyse data from their student records. Full anonymity was therefore not possible. This was explained in the research information sheet and the intern consent form included permission to collect data about them from the other three groups of respondents. Each respondent also signed consent. Issues of confidentiality were dealt with through the use of a careful coding system designed to protect the identity of individual respondents. A unique code number was allocated to each respondent and the database was kept off campus on an external notebook computer. All analyses, reports and articles have used only grouped or coded data and open ended comments were checked to eliminate any identifying features such as the names of hospitals, other staff members or colleagues, wards and structures.

3.9 RESEARCH BIAS AND LIMITATIONS TO THE STUDY

3.9.1 The study populations

This study was restricted to two groups of interns from a single medical school. The interns in each group had undergone very different undergraduate curricula. The results will therefore be generalisable only to the other interns in the graduation groups represented by the samples. However, other medical schools which are undertaking curriculum reform and facing similar challenges may find the information contained in this thesis of interest and assistance.

The two study groups were so close to each other that there exists the possibility of 'contamination' between them. Also, both were atypical groups in that each had something to prove. The graduates of the outgoing curriculum were to have been the first group to commence the new Graduate Entry Medical Programme, but the infrastructure and funding were not finalized in time and so they remained on the traditional curriculum. Many were

greatly relieved but a few were disappointed. There was a feeling amongst this final group of undergraduates that their needs were somewhat neglected as the Faculty prepared for the GEMP.

The group which did start the GEMP was very aware of the fact that they were in some way special and that they, and the staff, were exploring new methods such as problem based learning together. Much attention was given to explaining the rationale of the curriculum to students and they were offered several avenues of communication with the curriculum developers for feedback and suggestions, including scheduled meetings with the CHSE staff and online communication such as an electronic notice board for daily notices as well as an online feedback service to deal speedily with student concerns and to send a report on the resolution of the issue directly to the students' personal email addresses. These mechanisms for swift communication were not available to the undergraduates of the traditional curriculum. However, they had had the advantage of being able to 'learn the ropes' from previous generations of students which was not possible for the GEMP group.

Another difference was that the GEMP curriculum resources were delivered online with easy access to computers so that all aspects of the curriculum such as timetables, instructions and information, the lecturers' slide presentations, additional notes, skills checklists, resources, and images remained available throughout their undergraduate years. This must surely have provided this group with a much clearer understanding of the progression of the body system blocks and the vertical themes that ran through the four years of the GEMP, thus giving the curriculum a coherence and direction not available to previous groups of medical students.

It is inevitable that there must have been some Hawthorne effect (Cohen *et al*, 2007), where students were aware throughout their undergraduate training that they were the subject of much attention. This might have threatened the validity of the study to some extent. However, this "special case" effect pertained to both groups, though probably more so to the GEMP group. Another feature of the study mitigating against the Hawthorne Effect was that the data collection was retrospective rather than observational, so that it would not have had an influence on the interns' performance prior to the data collection.

3.9.2 *Research timing and method*

The new Graduate Entry Medical Programme commenced its clinical years (GEMP III) in 2005. This entailed a series of workshops to orientate clinical tutors (consultants and registrars) to the GEMP and to clinical teaching in a problem based learning curriculum. Since the same clinical tutors were still teaching Wits undergraduate students in MBChB VI of the old curriculum at that time there is the possibility that there could have been some carry-over effect of these new insights. These sixth year students became the 2006 intern study population. It was unfortunately not possible to start the research with the 2005 interns which would have partly avoided this source of bias.

It is not known to what extent undergraduate students in the GEMP communicated ideas and learning skills (e.g. online access to information and to the GEMP curriculum material itself) to students on the traditional curriculum.

A further limitation was the short time available to the researcher on the visits to each of the hospitals. The two interviewers had to maintain a very tight schedule between regular work commitments, with return flights and accommodation booked ahead of time and limited finances. Interview times were pre-arranged with the interns, and sometimes with supervisors but it was sometimes difficult to keep to the appointment time due to the fluctuating situation in the wards. Also, where no suitable patient or colleague was available during the visit, they had to be left out of the study. It was not possible to book interviews ahead of time with these two groups because of changing nursing duty rosters or patients having to go for diagnostic studies, treatments, having visitors or having been discharged from the hospital. This was less often a problem with the supervisors and it was usually possible to arrange an interview, sometimes after hours or telephonically.

Finally, the cross-sectional nature of the study meant that some of the intern competencies or behaviours had not been observed by the supervisors and colleagues by the time that the questionnaire administration and interviews were conducted. It became evident from the results that internship was not the best time to measure some of the desired outcomes of the

model of the competent South African intern. These included some important factors in the original decision for curriculum renewal. The first few months of internship had been selected because it was reasoned that at this stage the influence of the undergraduate curriculum could still be distinguished from learning gained during the internship itself. However, this unavoidably resulted in more missing data than was desirable and may have affected the significance of some of the results.

There is the ever-present possibility that the influence of medical undergraduate education is short lived with the introduction of numerous other influences during internship. These would include the hidden curriculum, role models, differing opportunities to learn and the influence of overseas doctors in the Health Services. For this reason the interviews were conducted as early as possible in the internship year

3.9.3 Instruments

Instruments that would have given a more objective measure of intern competence, such as having a supervisor rate the accuracy and detail of the interns' written bed notes, an OSCE and even perhaps a written theory examination were not feasible given the very widespread location of the interns. These would have added a valuable dimension to the study.

3.9.4 Respondents

Each group of respondents in the study might have introduced certain biases as indicated earlier in this chapter in Figure 3.5. These include how self-critical interns might be or the insight that they have regarding their own abilities as well as differences in culture, personality and upbringing. With supervisors there was the possibility of prejudice towards certain medical schools, personality clashes or halo effect introduced by giving responses about interns in general rather than the specific intern that they were asked to assess. Nurse colleagues might tend to evaluate performance according to how helpful the intern was to the nursing staff whereas for colleagues who were intern peers there was the possibility of rivalry and the fact that they were not always on call together. Patient issues included their sometimes

rather limited knowledge of the particular intern as they are seen by many doctors and the degree of discomfort that they were experiencing due to illness. The risk of "politeness bias" (Flynn, 2006) was possible for all respondents.

Respondent selection by the researcher might have introduced some bias in the case of both colleague and patient respondents. The intern suggested two or three colleagues and patients with whom he or she had worked closely enough to be able to respond to the interview schedules. The researcher then made the final selection based upon which of the colleagues was available at the time of the interview and which patients were in the ward and feeling well enough to be interviewed.

The potential problem of respondent's language did not arise and all of the patients interviewed were able to communicate satisfactorily in either English or Afrikaans, both languages being spoken by the researcher. On one or two occasions a nurse was asked interpret a specific question if the patient did not seem to fully understand.

3.9.5 Researcher

Judith Bell (1987) highlighted certain advantages to being an "inside researcher" and many of these benefits were true of this study such as a knowledge of the context of the research, the history of the University's curriculum change and the micro-politics of some of the institutions visited, access to people and documentation and an understanding of the hospital services and the problems of staff and students. However, Bell (*ibid*) also warned of issues which can be problematic. The researcher in this study had been involved in the GEMP curriculum from the outset and could have an inherent interest in a positive outcome for the research study thus introducing bias in the form of a lack of researcher objectivity. These warnings were heeded and a highly experienced, neutral and unbiased interviewer was employed for all interviews with interns and their supervisors in order to control for this source of bias.

3.9.6 *Potential for conflict of interest*

The current research was commissioned and funded by the institution which took the decision to implement curricular change. This required that the researcher remain constantly aware that this study should “aspire to value neutrality” and “make a contribution to the field” (Cohen *et al*, 2007, p.42) in order to be considered research rather than purely the evaluation of the new curriculum.

3.9.7 *The use of quotations in qualitative research*

The selection of individual comments in reporting qualitative research might be seen as giving a particular emphasis, thus impacting on the interpretation of results. For this reason, counts of comments in table form are included to indicate the strength of a particular theme. Individual comments remain just that, individual, and the selected comments are chosen as examples of the respondents’ voices in detailing issues relating to that theme.

3.10 SUMMARY

This chapter has dealt in detail with the design, methods and procedures of the study. A clear understanding of this is necessary in order to fully appreciate the complexity of the results which are presented in Chapter 4 and discussed in more detail in Chapter 5.