Chapter 3

Research methodology

Introduction

This study was designed to address the research questions outlined in Chapter 2, by investigating the approaches to learning, learning outcomes and the relationship between these two aspects of learning. The research focuses on the learning of a group of first–year science students studying a particular science topic, which emphasised the important biological concept of meiosis.

This chapter provides a description, justification and evaluation of the study methodology used to investigate the research questions. The context of the study is first outlined, including the biological concept used to measure students' learning outcomes. The research design is then described, including an overview of the data collection procedures, recruitment of participants and sample sizes. The research instruments of interviews, questionnaires and test/examination questions are then presented in more detail. The analytic procedures used are described, and finally the methodology is critically evaluated.

Context of the research

This section describes the specific group of students involved in this study, providing sufficient detail to enhance validity by enabling comparisons with other learning contexts. The biological concept of meiosis is explained, and some research into students' understanding of meiosis is outlined.

The broad context of this research was a first–year introductory biology unit, at the University of New England (UNE) in the regional city of Armidale in New South Wales. This is a relatively small, regional university, with enrolments of about 18,000 students, the vast majority of whom are from Anglo–Celtic, English–speaking backgrounds. About 40% of UNE's students are from rural and isolated areas, and over 50% are from the low socio-economic equity group. These figures are extremely high in comparison to state and sector averages, for example UNE's

43.7% rural access rate is nearly three times higher than the national rate of 16.3 % (UNE Planning and Institutional Research, 2005). In addition, about 70% of the student population studies off–campus. Many of UNE's students, therefore, have a different educational and social "capital" to that of more traditional university students, and the "gulf" between school and university for students such as these has been widely discussed (McInnes, James, & McNaught, 1995, p. 29).

The study was conducted within the second semester unit, Biology II (BIOL 120). Students from a range of degree programs undertake this unit, which is prerequisite to later units in the Bachelors of Science, Agriculture, Rural Science, Natural Resource Management and Environmental Science, as well as combined degree programs. The unit can be taken in either external or internal mode.

On–campus students enrolled in the internal mode attend three lectures and one of three repeated three–hour laboratory sessions per week. Off–campus, externally enrolled students receive printed resources covering the same content as the internal students, and attend a five–day intensive residential school, where the unit content is addressed in two lecture/tutorial sessions and two laboratory sessions each day.

The assessment for internally and externally enrolled students is slightly different. Both groups of students sit the same end of semester examination worth 50%. They also submit an assignment, which is worth 15% for the internal and 25% for external students. The internal students sit three practical tests worth 35% in total, instead of which the external students submit a report, based on residential school practical work, worth 25%.

Overall, this context was chosen for a combination of theoretical and practical reasons. The first year of university has been the focus of many studies, including investigations into learning approaches and outcomes described in the previous chapter. The research focus on first–year reflects the crucial importance of the first–year experience to students and tertiary institutions (described by McInnes, James, & Hartley, 2000; Pitkethly & Prosser, 2001), particularly given the diverse student body in this era of mass tertiary education (James, McInnis, & Devlin, 2002). The other more pragmatic reason for this choice of context was that the researcher had taught in this unit for some years, ending about two years prior to the commencement

of the project, so was familiar with the curriculum and the scientific concepts being taught. During the study, the researcher was employed elsewhere in the university, so the chosen study context maximised accessibility, knowledge of scientific content, and theoretical relevance to current literature.

Specific study context

The more specific context within which this research was conducted was a short topic on *Cellular and Organismal Reproduction* that featured the central and recurring concept of meiosis. This topic was taught by a single lecturer over two weeks at the beginning of second semester for internal students, and as part of a five–day residential school for external students. The topic discussed the biological concepts of cell division and reproduction in relation to organismal reproduction. The learning approaches students adopted within this topic were investigated, and students' learning outcomes were measured by assessing the variation in their understandings of the central concept of meiosis.

The choice of this specific context was, again, based on a number of theoretical and practical considerations. Questions relating to learning approaches and learning outcomes can be, and have been, addressed at a variety of scales. Coarser scale studies have examined learning approaches and/or outcomes across units of one year in length (e.g., Trigwell & Prosser, 1991), while finer scale studies have been conducted at the level of particular topics taught over a few weeks (e.g., Hazel, Prosser, & Trigwell, 2002; Prosser, Trigwell, Hazel, & Gallagher, 1994).

This study focused at the finer scale end of the spectrum. Previous studies in a first–year science context have shown that students use different learning approaches at different stages of a course (Millar, Prosser, & Sefton, 1989). This suggests that focusing on only one stage could provide a more precise picture of learning approaches adopted.

Likewise, many factors suggested that an in-depth investigation of a single concept was the most appropriate scale at which to assess variation in learning outcomes and relate outcomes to the learning approaches adopted. Examining qualitative variation in learning outcomes using SOLO (or other qualitative assessment tools such as phenomenography) entails detailed investigation of a manageable and reasonably

narrow conceptual content base. This tends to restrict the breadth of material that can be covered, particularly in the case of biological concepts that are composed of many interlinked subordinate concepts, and are actually quite broad when "unpacked". In addition, assessing the quality of learning outcomes using the SOLO model requires coding of students' written responses to carefully worded open–ended questions. The coding process is intellectually demanding and time–consuming, and past experience of others had suggested that "the broader the question the harder it was to code" (comment from teacher using SOLO, cited in Pegg & Tall, 2002, p. 253).

It is possibly for some of these reasons that most recent research on students' understanding of science using the SOLO model has focused at the level of specific concepts, for example, understandings of diffusion and osmosis (Panizzon & Pegg, 1997), photosynthesis (Hazel et al., 2002; Levins & Pegg, 1993) and natural selection (Creedy, 1993). Much other fruitful research into students' learning of science, for example, studies in alternative conceptions and expert — novice studies, has used a particular scientific concept as the unit of investigation. Moreover, an understanding of alternative student constructions/conceptions of specific scientific concepts such as meiosis is valuable in its own right, with the potential power to inform and improve teaching and curriculum decisions.

> By gaining knowledge about how students comprehend, for instance, various scientific principles and ideas, we should obtain information which would undoubtedly prove fruitful for teaching. (Marton & Säljö, 1976, p. 10)

This research focus on students learning approaches and outcomes over normal activities during a single topic in one unit at one university has a deliberate local emphasis on "a focused and bounded phenomenon embedded in its own context" (Miles & Huberman, 1994, p. 10). As argued by Miles and Huberman, a focus on events occurring in natural settings provides a good picture of what "real life" is like for these students. To address the research questions it was necessary that the study be based as much as possible on the authentic learning context of students. This meant observing, interpreting and describing students' approaches to learning and learning outcomes over the routine course of activities associated with the particular topic of the unit, or at least in close proximity to them (Miles & Huberman, 1994, p. 10).

Conceptual context for the study: meiosis

The conceptual context for the investigation of qualitative differences in learning outcomes was meiosis, an important and core component of the topic. This subsection presents some contextual information about meiosis as essential background to the current study.

Several advantages were associated with using meiosis as the context within which to assess learning outcomes. The concept had been introduced to the students several months earlier in the genetics component of the preceding first semester unit. The advantage of this was a potential levelling of some of the more extreme differences in prior understanding. Before enrolling in the first semester unit, some students had not studied science for many years, while some recent school leavers would have been familiar with the concept from high school biology. Although differences in understanding no doubt persisted, all the students at least had an equivalent recent exposure to the concept prior to this study

In addition, meiosis is an important biological concept that lies at the intersection of the major biological sub-disciplines of cell theory, evolution, genetics and reproduction. It is important in biology for several reasons. Meiosis enables cells and organisms to reproduce while maintaining (with some significant exceptions) one intact copy of the full genome, it provides the genetic basis for inheritance of characteristics and it contributes to genetic diversity, which is the raw material of natural selection and evolution (Knox, Ladiges, & Evans, 1994, pp. 147-151).

In simple terms, meiosis is a process that occurs in some cells of living things other than bacteria and related singlecelled organisms. It is a type of division of cell nuclei that results in four genetically unique daughter nuclei. Each of the daughter nuclei has half the number of chromosomes as the single parent nucleus and forms (in animals) the gametes, or eggs and sperm (Knox et al., 1994, p. 149). Meiosis is an intricate and complicated procedure, which is conventionally broken down into a number of phases or stages to describe the movements of chromosomes and other subcellular structures. These sub-processes collectively explain how one parent nucleus divides into four unique daughter nuclei with half the number of chromosomes, while maintaining the full genome.

The process of meiosis is described in full in Appendix A, including the names of the stages and phases, the movements of the chromosomes and the roles of subcellular structures involved in meiosis. Because it involves movement of microscopic subcellular structures, the process of meiosis cannot be seen directly by students using conventional compound microscope techniques.

Meiosis, along with its related subordinate concepts, is recognised as a difficult yet important concept for students to learn, and therefore has received considerable research attention (e.g., Brown, 1990; Finley, Stewart, & Yarroch, 1982; Kindfield, 1991; Longden, 1982; Smith, 1991). One study found that meiosis and the closely related mitosis were perceived by high school teachers as the third most difficult and second most important to learn in biology (Finley et al., 1982). Similarly, Johnstone and Mahmoud (1980) found that genetics, including concepts of DNA, cell structure and division, gametes and genes, was considered by high school students (and examiners) to be one of the four most difficult areas. Meiosis was identified by Hackling and Treagust (1982) as a poorly understood concept which limits high school students' understanding of inheritance.

A major contributor to students' difficulties with meiosis is thought to be the large number of subordinate concepts that must be mastered and brought together to understand the process (Smith, 1991). Students need to understand what DNA, genes and chromosomes are, and how these are related. In addition, they need to understand the sequence of meiotic events and stages that involve the chromosomes, and how these lead to the products of meiosis.

Summary

This section has outlined the practical and theoretical context of this research. It was conducted in a regional university with large numbers of external students, and students from rural areas and/or low socio-economic backgrounds. The more specific context was a topic within a first–year introductory biology unit, which was studied by both internal and external students. The research was conducted as far as possible within the normal activities and assessment tasks of the topic. Students' learning approaches and outcomes were investigated in relation to that topic, with students' understandings of the concept of meiosis providing a measure of their learning outcomes.

Research design

This section begins with a detailed, chronological outline of the data collection procedures as applied to the learning context, approaches and outcomes. This is followed by an explanation of recruitment procedures and the study sample sizes.

Data collection procedures

In accordance with the postpositivist framework of this research outlined in Chapter 1, the research design of this study has triangulated both quantitative and qualitative methods of data collection. As pointed out by Webb (1966, p. 35, cited in Denzin, 1989, p. 244), different methods of collecting data, for example, questionnaires, interviews and observation, have their own specific biases and validity threats, leading to the desirability of converging multiple research methods. This use of multiple methodological perspectives is known as triangulation (e.g., Denzin, 1989) or "critical multiplism" in terms of Lincoln and Guba (1994, p. 110).

The concept of triangulation has been criticized in some quarters for its positivist assumptions (reviewed by Denzin, 1989, pp. 244-246), and for the assumption that combining data from different methods adds up to a more complete picture (Silverman, 2001, p. 235). Nonetheless, triangulation is accepted by many researchers from a range of paradigms as "a strategy that adds rigor, breadth and depth to any investigation" (Denzin & Lincoln, 2003, p. 4), and is consistent with and characteristic of postpositivist inquiry (Guba & Lincoln, 1994, p. 110). As outlined in Chapter 1, harnessing complementary methods from different research paradigms has long been embraced in research into learning in higher education.

The three separate arms of the data collection were; first, observing and documenting the learning context; second, establishing the approaches that students took to learning in the topic; and third, gaining an understanding of the qualitative variation in their learning outcomes from that topic. A detailed outline of the procedures followed for these three aspects of the study is provided in Table 3.1.

Table 3.1: Summary details of data collection procedures

Activity	Research theme & line of inquiry	Resources and instruments	Year 1 schedule	Year 2 schedule
Relevant unit materials examined	1a	Biological Sciences 102 Unit Handbooks, textbook		
Internal lectures for topic attended and audiotaped	1a	Audio equipment	Wk 1: 24, 25, 27 July	Wk 1: 23, 24, 26 July
Internal laboratory session for topic observed and video/ audiotaped	1a	Audiovisual equipment Yr 1 (sessions observed but not taped in Yr 2 as the tapes from Yr 1 yielded very little information)	Wk 2: 30, 31 July, 2 Aug.	Wk 2: 29, 30 July, 1 Aug.
External lectures for topic attended and audiotaped and laboratory sessions observed	1a	Audio equipment	Residential school 22 Sept.	Residential school 27 Sept.
Focus group interviews conducted	1a, 1c	Semi-structured interview protocol, audio equipment	Internals 7, 8, 9 Aug. Externals 23 Sept.	Internals 6,7,8 Aug. Externals 27 Sept.
Questionnaires administered to internal students	1b, 1d, 3a, 3b	Two–factor SPQ modified for use at topic level: MSPQ Validity check questionnaire	Before practical tests returned in Wk 6	Before practical tests returned in Wk 6
Questionnaires administered to external students	1b, 1d, 3a, 3b	Two–factor SPQ modified for use at topic level: MSPQ Validity check questionnaire	During relevant section of residential school, 22 Sep.	During relevant section of residential school, 27 Sep.
Purposive subsample of individual students interviewed	1c, 1d, 2c, 2d, 3c	Semi-structured interview protocol, audio equipment	Not conducted in Yr 1	Internals: wk beg. 9 Sept. Externals: wk beg. 4 Oct.

Table 3.1 (cont.)

Activity	Research theme & line of inquiry	Resources and instruments	Year 1 schedule	Year 2 schedule
Target meiosis question prepared and included in internal practical test (N/A for externals)	2a, 2b, 2c, 2d, 3b	Question in practical tests	For practical tests in Wk 5	For practical tests in Wk 5
Target meiosis question (that was used in internal practical tests) given to external participants	2c, 2d, 3c	Question attached to MSPQ questionnaire. to inform sampling of external students for interviews	Not conducted in Yr 1	During relevant section of residential school, 27 Sep
Written responses of participants to practical test meiosis question photocopied	2a, 2b, 2c, 2d, 3b	Question in practical tests	Before practical tests returned in Wk 6	Before practical tests returned in Wk 6
Target meiosis question prepared and included in unit end-of-semester examination	2a, 2b, 2c, 2d, 3a, 3b	Examination question	For unit examination in Wk 14/15	For unit examination in Wk 14/15
Written responses of participants to exam meiosis question photocopied	2a, 2b, 2c, 2d, 3a, 3b	Examination question	After examination question marked by lecturer	After examination question marked by lecturer

As shown in Table 3.1, the data collection phase of the study took place over two consecutive years, with different cohorts of students participating in each year. The unit materials, assessment regime, practical activities and lecture content were the same across both years of the study. In each year, participants completed a questionnaire on their learning approaches (MSPQ), and their written responses to test and examination questions on meiosis were collected for analysis. In addition, individual interviews addressing both approaches and outcomes, and focus group interviews, were conducted with some students. All of these data were collected as closely as possible to the appropriate learning activities, within the constraints of the unit curriculum and students' schedules.

Recruitment of participants

The population of interest in this study was all students, both internally and externally enrolled, in the first–year unit Biology II at the University of New England. The population was sampled over two consecutive years in 2001 and 2002. The participants in this study were drawn from students enrolled in the unit, together with the relevant teaching staff. Lecturers and unit coordinators were approached prior to the commencement of the study, and were given an explanation of its aims and proposed methods. Their in-principle agreement was obtained, and the research proposal was then submitted to, and approved by, the university Human Research Ethics Committee (HREC), approval number HE01/150.

For both years of the study, the lecturers informed students of the project in the first lecture of the unit, and the researcher described the project to students in their first laboratory sessions (internal students) or at the residential school (external students). Information sheets and consent forms (Appendices B & C) were distributed to the students, and volunteers to be involved in the project were requested. In order to maximise participation and subsequent response rates for questionnaires, the researcher highlighted the central nature of student views and experiences to the project, and its potential contribution to the field of teaching and learning, and expressed her gratitude for any student involvement. From the pool of participants, subsamples of volunteers were later sought for exploratory focus groups interviews in both years, and individual interviews in Year 2.

Sample

Pertinent details of the sample are shown in Table 3.2. This describes the total sample for each component of the study, as well as details for Years 1 and 2.

Table 3.2: Research samples for both internal and external students

Aspect of research sample	Year 1			Year 2			Both years		
	Int	Ext	Total	Int	Ext	Total	Int	Ext	Total
Number of students enrolled in unit [1]	153	89	242	185	113	300	338	202	540
Participant numbers									
Practical test question	113	N/A	N/A	128	N/A	N/A	241	N/A	241
Exam question	118	58	176	132	26	158	250	84	334
MSPQ	87	50	137	138	27	165	225	77	302
Response rates (%)									
Practical test question	74	N/A	N/A	69	N/A	N/A	71	N/A	71
Exam question	77	65	73	71	23	53	74	38	62
MSPQ	57	56	57	75	24	55	66	38	56
Age (years):									
range	18-32	18-67	18-67	18-49	20-58	18-58	18-49	18-67	18-67
mean	19.9	29.4	23.1	20.1	33.3	22.0	20	31	22.5
median	19*	26*	20	19#	29#	19	19	27	19
Gender[2,3]:									
Females	68	40	108	67	16	83	135	56	191
Males	58	20	78	70	10	80	128	30	158
Degree[3]:									
Science	52	34	86	66	13	79	118	47	165
Rural Science	31	6	37	26	0	26	57	6	63
Agriculture	8	0	8	20	1	21	28	1	29
Nat. Resource	22	10	32	10	2	12	32	12	44
Envt'l Science	8	10	18	7	4	11	15	14	29
Arts	7	3	10	2	1	3	9	4	13
Other	0	0	0	3	5	8	3	5	8

Notes:
* Median age significantly different between internal and external samples, Year 1. Mann-Whitney *U* Test ($Z = -8.869$, $p. = .000$)
# Median age significantly different between internal and external samples, Year 2 Mann-Whitney *U* Test ($Z = -7.526$, $p. = .000$)
[1] Enrolment numbers according to unit administration records at census date.
[2] No significant relationship between sex and enrolment type (internal/external) in either year
[3] Some participants responded anonymously, so totals within sex and degree do not add up to total participants.

As shown in Table 3.2, most participants responded to both the learning approaches questionnaire and the appropriate measures of outcome. Because of attrition from the unit, absence or other factors, however, some participants responded only to the practical test or examination questions, or only to the questionnaire on learning approaches, or only to the validity check questions. From this sample, in depth individual interviews were conducted with eleven internal students and five external students in Year 2, and 58 students were involved in a total of eight focus group interviews; four interviews in each of the two years of the study.

Detailed information on response rates in the sample is also outlined in Table 3.2. This shows that in Year 1 the response rates for internal and external students were quite similar, but that response rates of internal students were about three times that of the external students in Year 2. The reason for this difference is not clear, but the lower rate for external students in Year 2 may have occurred as a consequence of asking these students for a written response to a meiosis question at the same time as asking them to read and fill in the information and consent forms and the MSPQ (see Table 3.1). This was done in order to assist in sampling for interviews. The additional task may have made participation seem more time consuming or anxiety inducing for these students, and thus reduced the response rate. It may also be the case that other factors associated with the learning context but unrelated to the study may have contributed to this difference in response rates in the Year 2 external cohort.

Table 3.2 also shows the distribution of age, gender and degree program across internal and external groups. Because the students' ages were not normally distributed, the non-parametric Mann-Whitney U test was used to test the significance of age differences. As shown in Table 3.2, external participants were significantly older than internal students in both years. The relationship between gender and enrolment type was examined using a χ^2 test for independence, which detected no significant difference between frequencies of males and females across enrolment type.

The broad pattern of degree enrolment varied slightly across internal and external students. In both groups the most frequent degree represented was science, and similar proportions of internal and external students were enrolled in Natural

Resources and Environmental Science degrees. The main point of difference was in the area of Rural Science/Agriculture which was the second most common area of study for internal students, but was studied by relatively few external participants.

Summary

This section has detailed aspects of the research design. The research triangulated quantitative and qualitative methods, to collect data from participants over two consecutive years. Participants responded to a self-report questionnaire on their learning approaches, and their written responses to examination and practical test questions on meiosis were collected for analysis.

The sample was over 300 students, though full data sets were available for fewer students, and smaller subsamples of students participated in individual and focus group interviews. Except for a lower response rate from external students in the second year, the response rates for various elements of the data collection were between 53% and 77%. Externally enrolled students were significantly older than the internal cohort.

Research instruments

As indicated by Table 3.1, three main complementary research instruments were used to gain information across the three aspects of the design. These were interviews with students, a questionnaire designed to measure learning approaches in the context of a particular topic, and test items to measure qualitative variation in students' learning outcomes. The rationale and procedures associated with these three instruments are described in this section.

Rationale and conduct of interviews

This subsection outlines the procedures followed for the student interviews. The overall procedure common to all interviews is discussed first, followed by a description of focus group and individual interviews.

Common procedural aspects

The interviews for all students began by thanking the students for volunteering, raising the issue of confidentiality, asking permission to tape the interview, and briefly explaining the purpose of the interviews and why they had been selected. Because knowledge questions can be threatening to participants (Maykut &

Morehouse, 1994, p. 91), it was also emphasised to students that they should ask for the interview to stop at any time if they felt stressed by the questions to follow. The interviews then started with one or two short, non-threatening background questions, and lasted from about 45 minutes to an hour.

The interviews were audiotaped, because, as argued by Silverman (2001, p. 162), audiotapes preserve conversations and can be replayed and transcribed, and they therefore enhance reliability. Only first names were used during interviews, and tapes of individual interviews were labelled by a six letter code (the first three letters of the student's first name and the first three letters of their last name). Although audiotaping introduces a potentially threatening aspect to interviews, possibly constraining students' responses, this was considered to be less intrusive than note-taking.

All interviews were transcribed in two phases. In the first phase the tapes were given to an experienced transcriber who was employed to make a first rough draft verbatim transcription. The researcher then listened to all the tapes, checking and filling in the initial draft transcriptions as appropriate. The final transcribed versions were for the most part verbatim, but contained some omissions of extraneous conversations, which were annotated on the transcriptions.

Focus groups

The focus group interviews, especially in the first year, were exploratory in intent, and had multiple aims. Because the study had been tightly defined by the researcher using questions drawn explicitly from a pre-existing theoretical framework, it was considered important to gather emic impressions of the significant aspects of learning in that particular topic. Although the questions framing this study focused on the learning outcomes and approaches components of the 3P model, it was also considered important to explore the students' perceptions of the learning context.

In addition, it was also hoped that these interviews would help to trial the amount of direction needed to optimise information relevant to the study questions, without leading the discussion excessively. Finally the group interviews were used to gauge students' reactions to questions being considered for later use in individual interviews, to inform whether, or how these questions might later be asked. Because

of these aims, the group interview protocol (Appendix D) was deliberately broad in nature, loosely structured and fluid, with further discussion themes added over the course of the interviews.

In total, eight focus group interviews were held, with from 6 to 12 participants in each, with larger numbers in the external groups. In line with recommendations of Carey (1995, p. 489) multiple interviews were conducted for internal students to minimise possible effects of "group-think" or dominant members, and to maximise generalisability. For internal students, three interviews per year were scheduled, one from each repeat laboratory session. This schedule was based on availability of volunteers and the time constraints of both students and the researcher, rather than the criterion of saturated data commonly recommended (e.g., Carey, 1995, p. 490). Because of concentrations of students from particular degree programs on particular laboratory sessions, this schedule in fact represented a de facto purposive sample across degree programs. More than two interviews were not possible for external students, as time constraints and the intensive schedule of residential schools restricted the number of volunteers so that only one focus group per year was conducted.

The group interviews for internal students were scheduled for immediately after practical sessions, in a seminar room adjoining the laboratory, to maximise convenience for students. In the case of external students the interviews took place immediately after dinner at the residential college where the students were staying. Students were personally recruited at random in laboratory sessions, working where possible on a rough rule of 50% over-recruitment to try to attain the optimum number of seven to ten participants (St John, 1999, p. 423).

The rooms were set up so that all the participants sat around a table, with a multi-directional microphone placed in the middle of the table and the tape recorder on the floor next to the researcher. As shown in the interview protocol, the interviews began with the researcher providing a brief explanation of the broad aims of the project and asking for diversity of opinion rather than consensus. As suggested by St John (1999, p. 426), the researcher explained that confidentiality in group interviews could not be guaranteed, expressed a personal commitment to confidentiality and requested confidentiality amongst participants. The interview then proceeded.

Individual interviews

The rationale for the individual interviews was to gather detailed, rich, case–based data from the students' perspective. While questionnaires provide broad, potentially generalisable information amenable to statistical analysis, qualitative techniques such as interviews can enhance more comprehensive exploration of ideas and impart a much deeper understanding of phenomena (Silverman, 2001, p 32), particularly complex phenomena such as the activity of learning. In addition, interviews were conducted to contribute to the reliability and validity of the research study, by triangulating interview data with both the SOLO codes of written responses and questionnaire results. They were also used to examine in-depth some common misconceptions and misuse of terminology in relation to meiosis, that relevant literature (e.g., Kindfield, 1994) and preliminary coding of students' written responses suggested might warrant further exploration.

The sampling strategy for individual interviews was purposive, with the aim of covering maximum variation in students. As advocated by Miles and Huberman (1994, p. 34) a sampling matrix was drawn up to guide the choice of which students to interview. Questionnaire scores were divided into high deep, high surface or combination, and preliminary SOLO codes of students' responses to the meiosis question fell into three major groups, so the resulting 3 x 3 matrix indicated the range of variation and combinations in learning approaches and/or outcomes of students. Students were selected to represent all cells of the matrix, except for two that contained no cases. This sampling strategy therefore included the typical, theoretically expected combinations of approach and outcome, as well as extreme or theoretically improbable approach and/or outcome scores. Within this overarching strategy, attempts were also made to sample as much as possible across gender, enrolment type and age.

The final sample composition and size (16, consisting of 11 internal and 5 external students) resulted from a combination of factors. Generally, the constraints included the number of students from each cell in the matrix who were available and willing to be interviewed, and the need to recruit and conduct interviews in a short time while the topic was still relatively recent. Specifically, preliminary coding showed very few students in the first cycle of the concrete–symbolic mode, some of whom were not subsequently available or declined to be interviewed, so very few

individuals in this group are represented in interviews. The final sample size of 16 compares favourably with the figure of 15 (plus or minus 10) typical of interview studies, which Kvale (1996, pp. 102) suggested reflects the law of diminishing returns in combination with time and resource constraints.

The interview procedure differed in some practical aspects, between internal and external students. Selected internal students were recruited in laboratory sessions, and a mutually agreeable time was negotiated for face-to-face interviews. These were held in a seminar room that was immediately adjacent to the laboratory and familiar to the students. Notes were not taken by the researcher during the interviews with internal students, so that eye contact, active listening and rapport with the student could be maintained. After individual interviews, short notes were made about any points of particular interest.

The selected external students were recruited in the residential school, but interviewed by telephone after they returned home. This decision was made because previous experience of external students, the intensity of the residential school schedule, and the feedback from recruitment of the focus group participants suggested that the students would be too busy and mentally overloaded to engage in additional activities while on campus. The students who agreed to participate provided their home telephone numbers, and were contacted by phone in the week following the residential school. These interviews were conducted using a standard speaker-phone which allowed both researcher and student to be taped. Short notes were taken during these interviews, mostly as reminders to the researcher to follow up certain points or to get back to a question that had been missed in the conduct of the interview.

A semi-structured interview format was chosen as most appropriate to the nature of the research questions, and research framework of the study. As pointed out by Miles and Huberman (1994, p. 35), for studies framed by specific research questions and using multiple cases, using well–structured instrumentation is a logical decision that facilitates comparisons across cases. Semi-structured interview protocols were therefore developed following widely accepted guidelines (e.g., Kvale, 1996; Maykut & Morehouse, 1994, pp. 83-100) about establishing rapport, framing and sequencing questions, and focusing interviews on the research questions. The

protocols focused on the themes of learning approaches (research theme 1) and understanding of meiosis (research theme 2), with the aim of investigating the relationship between these factors (research theme 3). Because of the different nature of the internal and external interviews, two slightly different protocols were used, which are included as Appendix E.

The preliminary interview comments and questions were followed by a series of tightly structured questions relating to the students' understanding of meiosis. In this stage of the interview the internal students were given copies of their test responses to read, and asked to clarify any ambiguities and whether they would change or could extend their answer. The students were then shown diagrams of chromosomes, and asked a number of questions to probe their understanding. They were then asked if they could explain the process of meiosis again, while drawing a diagram to demonstrate their understanding.

As suggested by Patton (1990, cited in Maykut & Morehouse, 1994, pp. 95-96), detail–orientated, clarification and elaboration probes were used to explore students' understandings as fully as possible, while endeavouring to retain an interview interaction rather than a stressful interrogation. Students were also prompted towards the end of the interview process to extend their explanations further. For the phone interviews with external students, visual aids were not possible. Therefore, in the meiosis section of the interview, each student's written responses to the meiosis question was read out to them, and the student's understanding of meiosis was probed using similar questions as for the internal students.

The second stage of the interviews for both internal and external students related to the approaches to learning the students adopted in the relevant topic of the unit. This was less structured than the first stage, but the questions (see Appendix E) focused on establishing:

- student motivation (e.g., Prosser & Trigwell, 1999, p. 91): students were asked about their interest in the topic and its relevance to them, and whether they enjoyed that section of the unit. The aim was to establish whether motivation was intrinsic interest or instrumental/pragmatic concerns.
- learning intention (e.g., Kember, 1990, p. 343): students were asked whether they were aiming to remember (to reproduce) or understand.

- learning focus and strategies (e.g., Prosser & Trigwell, 1999, p. 91): the intention of these questions was to find out whether the students were focused on words/diagrams and or meaning and exactly what did when (or if) they studied.

At the end of the interviews the students were thanked again, and asked if they had any questions that they wanted to ask the researcher. All questions asked by the students were answered, including questions about aspects of meiosis that required lengthy explanation.

Rationale and administration of questionnaires

This subsection considers aspects of the methodology relating to the learning approaches questionnaire. It begins with a discussion of the advantages of self-report questionnaires, describes the choice of the particular questionnaire used in this study, and how it was administered to the participants.

Advantages of questionnaires

Questionnaires were used in this study, as in much other educational research, for a variety of reasons. Quantitative data obtained from carefully worded questionnaires can assist in ameliorating some of the problems that have been identified with research interviews. For example, it has been argued that interview techniques can lead to "self–fulfilling prophecies" which reify pre-existing ideas (Richardson, 1994a, p. 452), and that methodological problems, such as response effect, arise from the social nature of interviews (Fleming, 1986). Richardson (1994a, p. 452) argued, therefore, that the repeated recognition of the deep and surface dichotomy could be an artefact of the interview technique. In consequence of this possibility, he highlighted the value of using instruments such as questionnaires to minimise personal interaction between researchers and participants.

Other advantages of using questionnaires are that they are less resource intensive than interviews, and they contribute to the reliability of studies (Cohen & Manion, 1994, p. 283). They also provide the large datasets appropriate for statistical analysis, testing of hypotheses and generalising conclusions.

Choice of learning approaches questionnaire

Of the many instruments available to measure learning approaches, the most appropriate for this study was an unpublished two–factor version of Biggs' (1987) Study Process Questionnaire (SPQ). A number of reasons led to the decision to use this particular instrument. It had been modified from Biggs' (1987) three–factor SPQ, by retaining only the Deep and Surface scales, and by rewording the items so that they were related to a particular topic (Prosser et al., 1994). These modifications addressed some of the major objections to the SPQ raised in previous studies (e.g., Richardson, 2000), for instance the weak support for the achieving scale and the desirability for an instrument of this kind to be administered in the context of a particular learning situation. The questionnaire used in this study is hereafter referred to as the MSPQ (modified SPQ).

Moreover, the contexts in which the MSPQ has been used previously were very similar to the context of this research, that is, short topics of study in first–year science units at Australian universities. The questionnaire items "were appropriate for student study in a several lecture topic in first year university science courses" (Prosser et al., 1994, p. 306). According to Michael Prosser (personal communication, 4 March 2002) the MSPQ has been used in a number of previous studies into learning approaches of tertiary science students (e.g., Crawford, Gordon, Nicholas, & Prosser, 1998a, 1998b; Hazel et al., 2002; Trigwell, Prosser, & Waterhouse, 1999), including Australian Research Council–funded research at the topic level (Prosser et al., 1994, p. 306).

The MSPQ had not been published in full, however, a copy was provided to the researcher on request (Michael Prosser, personal communication, 20 August 2001). The wording of the items in this study was exactly the same as in the version supplied. The relevant section of the unit was highlighted at the top of the questionnaires, along with the details of relevant lectures and practicals for internal and external students. The internal and external versions of the questionnaire (identical except for the headings) are included as Appendix F.

Administration of questionnaires

The decision on the timing of the administration of the MSPQ had to take into account several practical aspects of the unit, as well as relevant theoretical issues.

For the internal students, the MSPQ was administered in the laboratory sessions in the week immediately following the relevant practical test. This decision was made because of the strong influence of assessment on students' approaches to studying (e.g., Ramsden, 1992), and because much of the studying of the topic would be likely to occur immediately preceding its assessment, as assessment is well known to drive students' allocation of time and effort (Rowntree, 1987). For external students, the questionnaire was administered during the residential school laboratory session after the relevant lectures and practical activities.

Permission was granted by the relevant lecturers to use about 25 minutes at the beginning of laboratory sessions for administering the questionnaire. The instructions were read out to students, where it was stressed that they were to respond to the items only in relation to the *Cellular and Organismal Reproduction* section of the unit. Students were reminded about the details of that topic, including lecture and practical activities and concepts. Participants completed the questionnaires independently, while non-participants read through their practical notes for the session. Completed questionnaires were collected by the researcher.

Participants also completed a four–item validity check questionnaire at the same time as the MSPQ. This consisted of four questions including whether students remembered the relevant topic, and if/how their approaches to learning in that topic were any different from the subsequent unit context. This check was conducted because of the possibility that despite written and spoken instructions emphasising that it was to be answered in the context of the relevant topic, some students may not have remembered the topic, or may have answered the items in the context of some other topic more at the forefront of their mind. In addition, this instrument also included two short questions asking about students' background in meiosis. This questionnaire is included as Appendix G. For the external students in Year 2 of the study, a question asking about students' understanding of meiosis was also given to the students, to inform later sampling for interviews.

Rationale and development of test and examination questions

Information about the learning outcomes of students was sought in written responses of participants to questions about meiosis in practical tests and exams. This was to enhance the validity of the study by maximising the authenticity of the learning

context, and to minimise any demands on participants that were additional to the normal requirements of the unit. This subsection describes and justifies the choice of questions used to investigate students' learning outcomes from the target topic.

Questions testing students' understanding of meiosis were usual in practical tests and examinations in the years prior to this research. In accordance with past practice in the unit, only one question relevant to meiosis per test/examination was used. While it may have been preferable to ask more questions, or exactly the same question in practical tests and later in the exam, it was considered important to adhere to standard assessment practice in the unit.

The wording of the test questions was important, as inappropriately worded questions could impose upper limits on SOLO levels, or deny students the opportunity to demonstrate the full range of their learning (Biggs & Collis, 1982, p. 177). Permission was therefore sought from the appropriate lecturers for the researcher to suggest the wording of questions, which were then approved by the lecturers and incorporated by them in the usual way into the tests and examinations. The wording was developed in consultation with two other researchers experienced in writing items for SOLO coding. The aim was to generate open questions which would provoke extended responses suitable for coding, while still being commensurate with examination and practical test norms. Two questions were developed, one for use in practical tests and a slightly longer one for use in the examinations, and these are shown in Table 3.3 below.

Table 3.3: Questions used in practical tests and examinations:

Question context	Item
Practical test	Describe your understanding of the process of meiosis as fully as you can.
Examination	Describe in full the process of meiosis, and why it is so important in the life cycle of plants and animals

In both years the examination question was a compulsory component of a multiple–part question on cellular reproduction, and followed another compulsory sub-question on the cell cycle. Participants' responses to the question were photocopied by the researcher after they were marked by teaching staff in the unit, together with the front page of scripts containing identifying information. This enabled the measures of learning outcome to be matched to responses on the learning approach

questionnaire after coding. It also provided the opportunity to assess the criterion validity of the SOLO model by examining the association between independently assigned marks and SOLO levels.

Summary

This section has described the research instruments used in the study, which were selected by virtue of their suitability to the research questions, the complementary data they obtained, and their previous support in the literature. The instruments included interviews, questionnaires and examination questions, and this section has outlined the advantages associated with these instruments and how they were administered.

Data analysis

In this section the procedures used to analyse the data are described in detail, to counter the problem of analytic non-transparency alluded to by Miles and Huberman (1994, p. 2). This discussion begins with the first theme of learning approaches, followed by the second theme of learning outcomes, and then the third theme of the relationship between approaches and outcomes.

Except where otherwise indicated, all quantitative analyses were conducted using the statistical software *SPSS® Version 11* and *12 for Windows*, with reference to Pallant (2001) and SPSS (2005). Qualitative data from the interviews were analysed in a manner consistent with the postpositivist framework of this research.

Theme 1: Learning approaches

The results pertaining to learning approaches were analysed using an array of techniques. This subsection begins with a brief list of the techniques used, then describes the quantitative techniques, the qualitative techniques, and the specific aspects of data analyses adopted for each of the lines of inquiry, in turn.

Overview of analyses for learning approaches

An outline of the analytic techniques used in this theme of the study is shown in Table 3.4.

Table 3.4: Outline of analytic techniques associated with learning approaches theme

Line of inquiry	Data source	Analytic techniques
1a: context characteristics	Unit materials	Objectives, content & assessment relating to topic examined
	Lecture transcripts	Brief content analysis & coding
	Observation notes	Brief coding & comparison across different practical sessions
	Focus group transcripts	Data reduction & coding method of Miles & Huberman (1994)
1b: approaches to learning from MSPQ	MSPQ responses	Factor analysis Cronbach's alpha Rasch rating scale analysis *t*-tests of deep & surface measures across enrolment type & year & η^2 effect sizes Correlation between deep & surface scales Banded deep & surface measures *k*-means cluster analyses Rasch DIF analysis: internal & external groups *t*-tests of deep & surface measures between internal & external groups & η^2 effect sizes Between–groups ANOVAs: age, gender, enrolment type & learning approaches plus post-hoc Tukey's HSD
1c: approaches to learning from interviews	Individual interview transcripts	Data reduction, coding & display method of Miles & Huberman (1994)
1d: MSPQ vs interview data	Surface & deep measures from MSPQ & interview	Tabulation & comparison

As shown in Table 3.4, several data sources were used and analysed in the investigation of students' learning contexts. Analyses of students' responses to the MSPQ spanned a range of quantitative techniques. Some of these, such as factor analysis are widely used and well known in research into student learning in higher education, but others such as Rasch measurement modelling and *k*-means cluster analysis are less widely used. Analyses of interview transcripts followed the method outlined by Miles & Huberman (1994).

Quantitative techniques

Quantitative studies of students' approaches to learning frequently use principal components or factor analysis to analyse students' responses to learning process questionnaires (many such studies have been reviewed by Richardson, 1994a; 2000; 1999; Watkins, 1998). Factor analytic techniques explore the relationships between questionnaire items, by identifying a limited number of subgroups of correlated items, or factors (Tabachnick & Fidell, 1996, p. 635). In learning approach

questionnaires the ideal is that the empirically obtained factors correspond to theoretical constructs built into the questionnaire, in this case deep and surface approaches to learning. Responses to the items in each scale are then commonly totalled, resulting in deep and surface factor scores that may be used as interval–level data in parametric statistical techniques (e.g., Zeegers, 2001).

This practice, though, is complicated by the issue of whether Likert scales provide ordinal or interval measurement. Wiersma (1991, p. 177) described Likert scales as ordinal, but also stated that "the intervals between the points are assumed to be equal" (1991, p. 284). Similarly, Neuman (1994, pp . 198, 208-209), suggested that Likert scales are fundamentally ordinal, but they can be treated as interval–level measures, and that there is little statistical difference. By contrast, it has been argued that Likert scales are not intervals, and that "no conclusions can be drawn about the meaning of distances between scale positions" (Moser & Kalton, 1971, p. 364).

A second, related, criticism of adding factor scores discussed by Bond and Fox (2001, p. 66-68) is that questionnaire items for a given Likert scale, for instance deep learning approach, are likely to have different degrees of difficulty. For example, item 14 on the deep scale of the MSPQ ("I find that studying this topic is as interesting as a good novel or movie.") is a more extreme statement, that may be much harder for students to agree with, than item 10 ("I feel that this topic became interesting once I became involved in studying it."). Therefore, student agreement with item 14 should be weighted more highly than agreement with item 10 in gauging the extent to which students have adopted a deep approach to learning. By adding raw factor scores, though, the equivalent relative value of items is assumed.

Other disadvantages associated with using factor analysis in the context of student learning have been raised. Linacre (1998, p. 603) suggested that factor analysis can be misleading because non-linear data can generate pseudo–factors. Meyer and Shanahan (2003, p. 19) argued that by condensing data into a correlation matrix, factor analysis obscures individual and subgroup patterns of difference.

One alternative to factor analysis is Rasch measurement modelling, which has been strongly advocated as a more appropriate method of analysing responses to Likert scales (Bond & Fox, 2001, pp. 66-87). Although to date no investigations of learning

approach using versions of the SPQ have been analysed using a Rasch model, one study has used Rasch measurement in examining the Revised Approaches to Studying Inventory (RASI) (Waugh & Addison, 1998). The following subsection describes the principles and advantages of Rasch techniques in analysis of responses to Likert scales.

Rasch Measurement Modelling

The most appropriate of the Rasch family of measurement models to use in the analysis of responses to Likert instruments is the Rasch rating scale model. This approach transforms non-linear and non-interval raw scores from Likert scales to linear, interval–level scales. The five possible responses on the MSPQ (from 1 = Strongly Disagree through to 5 = Strongly Agree) are separated by four steps, or thresholds (from 1-2, 2-3, 3-4 & 4-5). The rating scale model estimates a relative difficulty for each of those thresholds and/or for the whole item by calculating the probability of students choosing each possible response. These "threshold values" and/or "item difficulties" are expressed in log odds (logits), on a linear scale. The probabilities take into account the relative "agreeabilities" of individual students, and "endorsabilities" of all the items (Bond & Fox, 2001, pp. 203-204; Wright & Masters, 1982, pp. 48-49; Wright & Mok, 2000, p. 100).

In addition, Rasch modelling has the advantage of the mathematical property of "specific objectivity" (Bond & Fox, 2001, pp. 146, 202-203), over using raw scores. Raw scores are directly tied to both the student sample and the specific items; if either of these is changed, the raw scores would change, and comparing raw scores across different instruments is problematic. Transforming raw item and student scores using Rasch techniques provides calibrated "measures" which are "freed of the particulars of the activities...[and] children [*i.e., items and students*] used to obtain them" (Wright & Masters, 1982, p. 34). In the context of this study this means that the representations of deep and surface approaches are independent of the specific items, and conversely, item difficulty estimations are independent of specific student samples.

The fit of the obtained responses to the rating scale model is indicated by infit and outfit mean square values. As argued by Bond and Fox (2001, p. 179), interpretation of fit statistics depends on the measurement context, and for Likert instruments infit

and outfit mean square values that fall between 0.6 and 1.4 are acceptable. These cut-off points correspond to 40% less and more variation respectively in the actual responses than the model would predict. In addition, infit and outfit mean square values that are transformed into a comparatively normalized *t* distribution provide infit *t* and outfit *t* statistics. These values are close to zero and have a standard deviation close to one when the actual data fit the model (Bond & Fox, 2001, p. 177).

The results of rating scale analysis provide information about the students and about the questionnaire items and scales. The extent to which students agree with the items on any particular scale provides a measure of their "ability" in the sense of "agree–ability" in the given context. The term "ability" here should not be taken to imply any intrinsic capability or predisposition to respond in a particular way. The ability measures are interval–level scores which are analogous in function to factor scores. For example, in the context of the deep and surface scales of the MSPQ, the deep and surface ability estimates for each student are a measure of the extent to which his or her responses agreed with deep and/or surface questionnaire items in the specified context.

The extent to which items are endorsed by students indicates the range of variation in item difficulties. Separability indices provide a measure of reliability on a scale of zero (low reliability) to one (high reliability) (Waugh & Addison, 1998, p. 103). These are conceptually similar to Cronbach's alpha reliability estimate, in that they estimate the proportion of observed variance considered to be true (Bond & Fox, 2001, p. 207). Fit statistics provide a measure of the validity of the deep and surface constructs of the questionnaire.

Rasch measurement using the rating scale model can also compare responses of subgroups to examine whether different groups of students perform differently on scale items. In Rasch terms, this is called Differential Item Functioning (DIF), which is analysed using the COMPARE command of the *Quest* software. This procedure estimates item difficulties for the two groups of students separately, and compares them (Bond & Fox, 2001, pp. 170-171).

The results from these analyses are presented in a range of outputs. Estimates of individual item difficulties and students' agreeabilities are calculated in logits.

Threshold values for each item and the performance of the students are commonly displayed on a map using the same item/person interval logit scale (e.g., Waugh & Addison, 1998). On these maps, thresholds or items endorsed by most students (easiest to endorse) are at the bottom, and thresholds or items endorsed less frequently (most difficult to endorse) are at the top. The number of students performing at each associated logit interval is also indicated, with students agreeing least (disagreeing most) at the bottom, and those with high agreeability estimates at the top. The fit statistics of the items are displayed on a separate item fit map that shows which, if any, of the items do not fit the model.

Rasch measurement techniques therefore have several advantages; they provide interval–level linear scales, measure the relative difficulty of items, and establish the relative abilities of students to endorse items taking into account the different item difficulties. In addition, they provide information about the reliability and validity of Likert instruments.

Unlike Rasch measurement techniques, which have been used relatively rarely in studies of students' approaches to learning, cluster analysis has been used in many recent studies of student learning (e.g., Crawford et al., 1998a; Hazel et al., 2002; Long, 2003; Meyer & Shanahan, 2003). It has been suggested that "an increase in the use of cluster analysis in research on student learning in higher education seems to be warranted" (Crawford et al., 1998a, p. 466). The following subsection provides a brief description of cluster analysis as a tool for research into tertiary student learning.

Cluster Analysis

Cluster analysis is a family of techniques used to detect and classify relatively homogeneous groups (clusters) within a data set (for a detailed recent review see Everitt, Landau, & Leese, 2001). Studies of student learning have used both hierarchical (e.g., Crawford et al., 1998a; Hazel et al., 2002) and non-hierarchical (e.g., Long, 2003; Meyer & Shanahan, 2003) clustering methods. Recent studies cited by Meyer and Shanahan (2003, p. 10) highlighted the effectiveness of the non-hierarchical *k*-means cluster analysis in investigating learning approaches in tertiary students.

The *k*-means cluster analysis is an exploratory technique that aims to maximise dissimilarity between groups, while minimising dissimilarity within groups (Meyer & Shanahan, 2003, p. 10), in identifying relatively homogeneous groups within the data. The technique requires that a number of clusters be specified *a priori*, and cases are then iteratively assigned to clusters based on the simple Euclidean distance from the cluster centres, until this process converges at a stable cluster solution. The final cluster centres are the means of each of the variables in the different clusters (Everitt et al., 2001, pp. 99-102; SPSS, 2005), and "reflect the characteristics of the typical case for each cluster" (SPSS, 2005). The analysis is repeated with increasing numbers of specified clusters, with the final choice of the best number of clusters resulting from comparisons of the different cluster solutions. Limitations of *k*-means clustering are that it is inappropriate for ordinal data and that clusters are scale dependent, requiring standardisation of variables that are measured on different scales (Everitt et al., 2001, p. 102).

Qualitative techniques

It has been argued that analysing patterns and relationships in lengthy text such as interview transcripts can "overload human's information–processing capabilities" (Faust, 1982, cited in Miles & Huberman, 1994, p. 11). This (by way of an irresistible digression) is in precise accordance with the principles of the SOLO model used in this research study. Because extended text is dispersed over many pages, sequential and not clearly ordered, comparisons of texts is extremely difficult.

It has therefore been suggested that distilled, organised displays of data are essential for valid analysis of lengthy text (Miles & Huberman, 1994). These researchers also suggested that communicating and reporting the results of such analyses should likewise use data matrices or networks to maximise the transparency of the conclusions to the reader. This argument is consistent with the critical realist ontology of its authors (Miles & Huberman, 1994, p. 4). However, it is counter to some more interpretivist research traditions, where rich, detailed narrative prose is often preferred for "thick" description and communication of research (Denzin & Lincoln, 2003, p. 16; Neuman, 1994, p. 449),

Because of the postpositivist framework of this research, together with its specific research questions and semi-structured interview protocols, interview analysis

followed in many respects the data reduction, coding and display procedures advocated by Miles and Huberman (1994). It was, however, decided that the scale was small enough and the interview data structured enough so that analysis of interviews could be carried out using the functions of a word-processor/spreadsheet rather than specific code-and-retrieve software.

The transcripts of individual interviews were examined numerous times, focusing solely on learning approaches initially, and then solely on learning outcome. These two major sections of the interviews were analysed independently of each other. Because the broad concepts of interest in individual interviews were clearly defined *a priori*, two conceptually clustered matrices were chosen as the most appropriate format for the analysis: one for learning approaches and the other for learning outcome. These matrices facilitate cross-case comparisons, enable a number or research questions to be considered simultaneously (Miles & Huberman, 1994, pp. 101, 127-131), and therefore assist the task of making meaning. The matrices were framed containing all students who participated in individual interviews, and theoretically important aspects of the research questions as headings.

For each student, the matrices were then filled in with four sorts of entries, as suggested by Miles and Huberman (1994, p. 129). Firstly, relevant sections of the transcripts relating to or illustrating some dimension of the headings were placed verbatim in the appropriate cell of the matrix. The entries were then examined and code labels were developed and entered for the patterns that emerged from this process. Some summary explanatory comments and ratings of some aspects of the data were added to facilitate comparisons.

To enhance reliability, the matrices of transcript selections and the labels derived from them were shown to two co-researchers, who scrutinised the appropriateness of the code labels in relation to the relevant transcript excerpts. In addition, reliability of the final coding was tested by showing the transcripts to an additional judge. This person was asked to examine the transcripts, without reference to the coding scheme, in order to provide an independent opinion on whether the transcripts indicated surface or deep approaches to learning. The entries in the matrices were then condensed to communicate summary findings from the interview process on students' approaches to learning.

Analysis of focus group interviews followed a similar pattern and rationale as the individual interviews. A matrix was framed containing the eight different focus groups and theoretically important aspects of the research questions as headings. Both patterns of similarity within and among focus groups, and diversity of responses, were sought.

Line of inquiry 1a: context of students' learning approaches

The unit materials were examined with particular focus on teaching and learning objectives, the content and sequence of instruction relevant to the topic in this study, and the assessment regime. A brief content analysis was conducted on the lecture transcripts, comparing content coverage across both years, and coding based on any themes that emerged during re-reading. Notes on observations of practical classes and lectures were reread several times, compared across years and internal/external groups, and emergent themes summarised.

The focus groups supplied crucial information on participants' perceptions of the learning context. In the case of the focus groups the headings in the analytic matrix were based on key aspects of the learning environment shown to be important to students' learning experiences (e.g., Prosser & Trigwell, 1999; Ramsden, 1979, 1992). These were:

- perceptions of teaching methods
- perceptions of the amount and difficulty of content
- perceived relevance and interest
- perceived purpose of tertiary study.

Rather than displaying the lengthy resulting matrix, major findings were summarised and illustrated by indicative excerpts from the matrix.

Line of inquiry 1b: Students' learning approaches from MSPQ responses

Analysis of MSPQ responses first focuses on the broad association between the deep and surface variables. The focus is then turned onto the variations in patterns of relationship between deep and surface learning approaches.

After checking the data for errors, questionnaire responses were explored using a number of Principal Components Analyses with Varimax rotation (assuming independence of factors) and direct quartimin oblique rotation (allowing correlations

among factors), as recommended by Tabachnick & Fidell (1996, pp. 666-668). This allowed comparison of the factor structure with other studies. These analyses were conducted on the entire data set, that is, both years' responses from external and internal students combined, as well as explored separately for internal and external students, and Year 1 and 2 of the study. Catell's scree test (Catell, 1966, cited in Pallant, 2001, p. 154) was used to determine the number of factors to extract, with all factors above an "elbow" in the plot being retained. Following the factor analyses, scale reliabilities were assessed using Cronbach's alpha.

Students' responses to the MSPQ were also examined using the computer software *Quest* (Adams & Khoo, 1993) to run the Rasch rating scale model. This was in part to establish measures of students' learning approaches as an alternative to raw factor scores, and to provide further evidence on the validity and reliability of the deep and surface scales. All 28 items were analysed as a single overarching "Approaches to Learning" scale, comprising the four subscales, and this provided measures of deep, surface, motive and strategy agree-abilities of the students, that is, scale scores. In addition, the separate deep, surface, motive and strategy subscales were analysed as independent scales.

The results of the five separate analyses were examined to check whether the scale/s satisfied the unidimensionality requirement of the model. The fit statistics of the single scale and the four separate subscales supported the fit of the data to the model. Therefore, the deep (motive and strategy) and surface (motive and strategy) case estimates from the analyses were adopted as measures of the students' deep and surface learning – that is, learning approach "abilities".

The ability measures of deep and surface approaches to learning from the rating scale analysis were then used in a variety of statistical techniques to examine different aspects of the students' learning approaches, in terms of relationships between the different variables. The data were initially tested for evidence of non-normality, and to look for evidence of bias resulting from different response rates for the two sampling years, the homogeneity of the Year 1 and 2 samples with respect to learning approaches was explored. The mean deep and surface ability measures were compared for each year, for both internal and external students, using independent samples *t*-tests.

Effect sizes for independent samples *t*-tests were calculated using the formula $\eta^2 = t^2/t^2 + (N_1 + N_2 - 2)$. These were interpreted according to Cohen 1988 (cited in Pallant, 2001, p. 184), that is, $\eta^2 = 0.01$ is a small effect, $\eta^2 = 0.06$ a moderate effect, and $\eta^2 = 0.14$ a large effect. This enables an assessment of the meaningfulness of any statistically significant differences in student learning approach, as advocated by Richardson (2004, p. 352). This was particularly important given the unequal group sizes being compared. To examine any relationship between deep and surface approach abilities, a Pearson product-moment correlation of these variables was conducted for the whole sample, and for internal and external students separately.

The patterns of relationship between deep and surface learning approaches in the entire sample were then explored. Scores on each of the deep and surface measures were grouped into five bands, with the top 10% of scores placed in the highest band for each scale, the next 20% in the next highest band, the middle 40% in a medium band, the next 20% and the last 10% in two lower bands. A matrix of each combination of deep and surface ability bands was constructed to show the extent of varying combinations of learning approach.

The patterns of relationships between deep and surface approaches at a finer scale were examined using by *k*-means cluster analyses. For these analyses the deep and surface scales were each subdivided into their motive and strategy components, and the deep motive, deep strategy, surface motive and surface strategy learning approach abilities converted to *z*-scores. Then cluster analyses of the four learning approach subscales were conducted sequentially, specifying from two through to ten clusters with each repetition. The cluster mean scores for each of the analyses were graphed, to aid in identifying patterns and differences in successive solutions. The different cluster solutions were then compared with the aim of locating two successive solutions that had relatively stable cluster size and membership or recognisable variations thereof. This method for selecting appropriate cluster solutions closely follows that used by Meyer and Shanahan (2003).

In addition to these analyses, the variables of student age and gender were examined to test for relationships with learning approach. Because the distribution of students' ages was markedly non-normal, the students were divided into three age-groups (18-19, 20-21 & 22+) for a number of age comparisons. Two–way between–groups

ANOVAs were used to test the effect of age and gender on the deep and surface ability scores established by the Rasch measurement modelling. Cluster membership of each student was also used as a basis on which to compare age and gender composition of the different cluster groups.

Comparison of internal and external students

The student sample in this study comprised the two different subgroups of internally and externally enrolled students. Line of inquiry 1b hypothesised that these two groups would exhibit differences in learning approaches given the many differences between them. This question was addressed using a number of methods.

Correlations were calculated for internal and external groups separately, and independent samples *t*-tests were conducted to compare the means of the deep and surface agreeability estimates for internal and external students. The ratio of internal:external students in the different deep and surface ability bands was calculated. In addition, the agreeability estimates for the two subgroups were compared using a Rasch Differential Item Functioning (DIF) analysis, to examine the performance of the two groups of students on the individual deep and surface items of the MSPQ. Cluster analyses of internal and external subgroups were also conducted separately, following the procedure outlined for the full sample, but partitioning the sample into internal and external students.

Line of inquiry 1c: Analysis of interview data on learning approaches

For the learning approaches aspect of the study, the headings for the interview analyses were based on key indicators of learning approach as reported in the literature. These headings and associated indicators used in the analysis were:

- student motivation (e.g., Prosser & Trigwell, 1999, p. 91). Satisfying intrinsic interest was used as an indicator of a deep approach and instrumental or pragmatic motivation was indicative of a surface approach.
- learning intention (e.g., Kember, 1990, p. 343). The two key indicators were the intention to reproduce (surface) or understand (deep). Because of the predominant importance of learning intention on learning approaches adopted, this indicator was weighted most heavily in assessing individual students' approaches to learning.

- learning focus (e.g., Prosser & Trigwell, 1999, p. 91). This included a range of study behaviours associated either with a focus on meaning, indicative of a deep approach, or a focus on words or diagrams indicative of a surface approach
- learning strategies (e.g., Meyer & Shanahan, 2003). Different types of memorisation identified from phenomenographic and subsequent psychometric studies (reviewed by Meyer & Shanahan, 2003) were also used as secondary indicators of approach. These had not been built into the interview protocol (which preceded publication of Meyer & Shanahan's paper) but were apparent during preliminary interview analysis and were therefore taken account of in the coding process. Memorising before understanding and as a rehearsal strategy indicated surface approaches to learning, while memorising after understanding, memorising with understanding and repetition for understanding were considered as indicative of deep approaches.

These headings and indicators were used as coding labels for assessing students' learning approaches. An explanatory table of the indicators for each of the four headings together with example extracts from interviews was developed, to enhance the transparency of the coding meaning. Interview data were displayed according to final determination of learning approach adopted, with indicative interview excerpts.

Line of inquiry 1d: Relationship between quantitative and qualitative data

The quantitative and qualitative data were compared in a summary table, as a check of the validity of both the quantitative and qualitative findings. For the quantitative data, the MSPQ deep and surface ability estimates, and the associated ability band for each scale, were tabulated for each student. Students were coded as taking a predominantly surface approach if their ability estimate was higher in the surface than the deep scale, and as taking a predominantly deep approach if their ability estimate was higher in the deep than the surface scale. Students in the same ability band for both scales were coded as showing mixed approaches to learning. These final determinations from the MPSQ responses were compared to the determinations from interviews.

In summary, this subsection has discussed the analyses of data collected for the first study theme of students' approaches to learning. Three quantitative techniques of factor analysis, Rasch measurement modelling and cluster analysis have

complementary uses and advantages that are particularly appropriate to analysis of MSPQ responses. The data display, coding and reduction technique of Miles and Huberman (1994) is consistent with the philosophical foundations of this research and was accordingly used to analyse interview transcripts relating to learning approaches.

Theme 2: Learning outcomes

The analyses pertaining to the theme of learning outcomes were based on categories of qualitatively different written and verbal explanations of meiosis, together with marks for written responses. This subsection begins with an overview of the analyses used in this study theme, followed by a description of the specific analyses used for each of the four lines of inquiry.

Overview of analyses for learning outcomes

An outline of the analytic techniques used in this theme of the study is shown in Table 3.5.

Table 3.5: Outline of analytic techniques associated with learning outcomes theme

Line of inquiry	Data source	Analytic techniques
2a: categories of written response & fit to SOLO	Written responses & marks to meiosis questions	Categorisation of responses with co-researchers Comparison of hierarchy of categories with SOLO model, inter-judge reliability check Spearman correlation of SOLO levels & marks
2b: distribution of SOLO categories	SOLO codes & marks for each written response	Frequency tables & histograms χ^2 tests of independence & goodness of fit *t*-test marks: internal & external cohorts & η^2 effect sizes Pooled Low ($\leq M_2$) & High ($\geq R_2$) groups Cross-tabulation of enrolment, age, gender & SOLO groups
2c: SOLO categories from interview	Individual interview transcripts	Coding according to SOLO levels obtained from written responses, using matrix technique (Miles & Huberman, 1994)
2d: written vs interview data	SOLO codes from written responses & interviews	Tabulation & comparison

As indicated in Table 3.5, the analyses for data relating to learning outcomes initially comprised a categorisation of written responses, then a comparison of those categories with the SOLO model. The resulting SOLO levels were used as variables in a range of standard non-parametric techniques. Students' verbal responses were also coded using SOLO, and written and verbal explanations were compared.

Line of inquiry 2a: Analysis of written responses to meiosis questions

The analysis of the qualitative differences in written responses to the meiosis questions required a number of steps. The responses to both practical tests and examination questions were initially marked in the usual way by the teaching staff of the unit. This provided an independent quantitative assessment of responses.

To examine qualitative variation in responses to items in practical tests and exams, the researcher read through the entire set of one year's responses, grouping like responses, refining and re-refining the emerging categories over many passes through the material. During this process, the researcher suspended her knowledge of the SOLO framework as much as possible. Many responses used diagrams, and during the analytic process the structure and meaning expressed in the diagrams as well as the text were used to categorise responses. For the examination question, coding was based only on the first part of the question: "Describe in full the process of meiosis". At each pass, emerging similarities and key points were summarised into a written record of progress.

Throughout this process, periodic discussions were held with two co-researchers experienced in coding SOLO responses, one of whom was almost totally unfamiliar with the content matter while the other had a good working knowledge of meiosis. In these discussions the co-researchers examined samples of the original responses, compared them with written summaries of progress and raised questions, objections and suggestions about the emerging categorisation of responses. In order to assess the variation in learning outcomes, the responses were examined, and patterns of variation in responses documented by the researcher.

The classification procedure continued until the categories of responses stabilised with re-reading. The researcher compared this hierarchy with the two-learning-cycle version of the SOLO model (see Chapter 2) and formed conclusions about the extent to which the hierarchy of groups reflected different levels in the SOLO model.

Written descriptions and a sample response for each of these groups were provided to the co-researchers, who also examined the hierarchy against the SOLO model and came to the same conclusions as the researcher. The final coding scheme was then established, and tested for inter-judge reliability by giving a pre-coded sample of 20

responses to the co-researchers, who coded them together using the scheme provided. The two sets of codes were then compared, which identified two responses on which there was disagreement, from two adjacent categories. The written description of these two categories was refined, and a further inter-judge test of responses from these two categories was conducted. Once the coding scheme was finalised, all responses were coded.

To test the criterion validity of the two-learning-cycle version of the SOLO model, Spearman's Rank Order Correlation was used to assess the relationship between practical test and examination SOLO codes and independently assigned marks.

Line of inquiry 2b: Patterns of distribution of SOLO categories

The qualitative analysis of the SOLO levels of students' responses to the meiosis questions provided a SOLO code for each participant, and a table of the relative frequencies of each SOLO level. The comparability of the learning outcomes for the Year 1 and 2 groups of the sample was assessed by χ^2 tests of the independence of year of study and SOLO mode for practical test question and the exam question. Categories were pooled into modes, as too many cells had an expected frequency of less than five at a category level, and the modes represented a theoretically logical unit of analysis.

The relative frequencies of SOLO levels in the sample were explored using χ^2 tests, and the categories were then pooled into two groups of low ($\leq M_2$) and high ($\geq R_2$) responses for later comparisons. Investigation of these two groups was of particular interest, because of previous research suggesting that surface approaches may limit outcomes to M_2 or below in the two-cycle model (see Chapter 2), and that deep approaches are associated with responses at or better than R_2.

The relationship between learning outcome, enrolment type (whether internal or external) age and gender were explored using cross-tabulations of these variables across the low and high SOLO groups. An independent samples *t*-test was also conducted to compare the marks attained in the examination question for external and internal students.

Line of inquiry 2c: Analysis of learning outcomes from interviews

The headings used to code the responses for the learning outcomes aspect of the study were based on the previous analysis of written responses, and included the different elements, the structure of the responses and their degree of abstraction. The structure of the responses was based on the level of conceptual integration the students demonstrated between the different aspects of the process of meiosis. Based on coding labels attached to these headings, students' verbal explanations of meiosis were coded accorded to the SOLO hierarchy established from the written responses.

Final data were presented by researcher summary comments only. This was because the underlying excerpts spanned many pages of close interconnected questioning/answering which could not readily be segmented for data display. Misconceptions and areas of confusion were looked for in later passes through the transcripts, by comparing the students' accounts of meiosis and its subordinate concepts with scientifically accepted understandings.

In summary, this subsection has outlined the analyses of data for the second research theme, which was based on students' written and verbal explanations of meiosis. The first line of inquiry was analysed by collecting independently assigned marks, categorising responses and comparing the categories with the two-learning-cycle version of SOLO. The relationship between marks and SOLO categories was explored. The second line of inquiry relating to distribution of SOLO categories was analysed by a range of statistical techniques, particularly χ^2 tests of frequency data. The data from interviews were collected for the third line of inquiry, and these were analysed by coding students' verbal explanations of meiosis according to the SOLO coding scheme established from the written responses.

Theme 3: Relationships between approaches and outcomes

This subsection describes the analyses used in the third research theme, which was to explore the relationship between learning approaches and outcomes. The initial hypothesis for this question was that surface approaches would be associated with responses to meiosis questions coded at M_2 or below, while deep learning approaches would be associated with responses to meiosis questions coded at R_2 or above. The first part of the subsection provides an overview of the analyses used

within this theme. The second and final part discusses the specific application of these analyses in exploring all the lines of inquiry within this theme.

Overview of analyses for the relationship between approaches and outcomes

An outline of the analytic techniques used in this theme of the study is shown in Table 3.6.

Table 3.6: Outline of analytic techniques associated with the theme of the relationship between approaches and outcomes

Line of inquiry	Data source	Analytic techniques
3a: relationship between MSPQ responses & outcome indicators	• SOLO codes from written responses • Surface & deep measures from MSPQ	- Spearman correlation between approach & outcome indicators - *t*-tests of mean deep & surface measures for High & Low SOLO groups & η^2 effect sizes - Cross-tabulation of deep & surface abilities with low & high SOLO levels - χ^2 tests of association between SOLO group & predominant approach - χ^2 tests of frequencies of High & Low SOLO groups in different learning approach clusters.
3b: internal vs external cohorts	• As for 3a	- As for 3a, but comparing internal & external groups
3c: approaches vs outcomes from interviews	• SOLO codes & approaches from transcripts	- Tabulation & comparison

Table 3.6 shows the three lines of inquiry, which were informed by data derived from analyses within the first and second study themes. The relationships between these data were tested using standard statistical techniques, and tabulation and comparison of qualitative data.

Lines of inquiry 3a-3c

As the SOLO categories established for the meiosis responses provided ordinal data, Spearman's Rank Order Correlation was used to investigate the broad overall relationships between deep and surface learning abilities, and SOLO categories and marks for the examination question responses. To assess the hypothesised relationship between approach and outcome, independentsamples *t*-tests were conducted to compare mean deep and surface learning approach abilities for the $\leq M_2$ and $\geq R_2$ groups, and η^2 values calculated to check effect sizes. In addition, the frequencies of students in different bands of surface and deep abilities were crosstabulated, and examined using χ^2 tests.

These analyses only inform the evaluation of relationships between learning outcome variables and deep or surface ability variables as isolated constructs. The relationship between outcomes and the more realistic combinations of deep and/or surface approach reported by the students in the MSPQ was also tested. Students were defined as showing a predominantly deep approach if they were in a higher deep than surface band, and showing a predominantly surface approach if they were in a higher surface than deep band. Those who were in the same band for each were defined as mixed. The frequencies of low and high SOLO responses were crosstabulated with these predominant approaches from the banded MSPQ responses, and differences examined using χ^2 tests.

At a finer–grained level, the clusters of students established on the basis of learning approach variables were used to explore the frequencies of low and high SOLO levels in different mixes of deep and surface approaches. To explore any differences between the relationships of learning approach and outcome in internally and externally enrolled students, the analyses outlined above were conducted separately for internal and external students.

The preceding analyses of the relationship between approach and outcomes were at the level of variables or groups of students with similar approaches. At the individual level, after several months of successive analysis of the approaches then outcomes sections of the interviews, results from the approaches and outcomes were compared. This was done by combining the summarised approach and outcome interview matrices, to examine any relationship between the learning approaches expressed in the interviews and the students' understanding of meiosis as articulated in the interviews.

In summary, analyses of the relationship between approach and outcome focused at three levels. The first was the relationship between approach and outcome variables, which was tested using correlations and comparisons of means. The second level was the relationship between approach and outcomes for groups of students showing similar patterns of learning approaches. This was analysed by χ^2 tests of frequency data using clusters and approach-band membership as categories. The third level was the individual student, and individual approach and outcome data were compared by

comparing summarised approach and outcome matrices from students' interview responses.

Summary

This section has presented the methods used in analysing the data collected during this study. The quantitative aspects of the analysis combined well–established techniques with others, like Rasch measurement modelling, that are relatively novel in the learning approaches context. The qualitative techniques used were chosen on the basis of their suitability to the study questions, the theoretically grounded nature of the study, and its postpositivist perspective. The analysis of learning outcome has utilised recent developments in the SOLO model for evaluating the quality of learning outcomes. Use of this range of analyses enabled integration of data within and between different aspects of the study, and thorough testing of the different research avenues being explored.

Evaluation of methodology

In this section the methodology used in the research study is subject to critical evaluation, in the spirit of Kvale's (1996, p. 242) suggestion of playing devil's advocate to one's own research. The section begins with discussion of reliability and validity. In the remainder of the section the criteria of reliability and validity are used to critically evaluate the study methodology. This evaluation focuses initially on the overall research design, followed by the interview–based aspects of the study. The reliability and validity of students' responses to the learning approaches questionnaire, and test and examination questions are then discussed. Finally, the ethical aspects of the research methodology are scrutinised.

Reliability and validity

Although interpretivist researchers argue for criteria such as trustworthiness and authenticity in evaluating qualitative research (Guba & Lincoln, 1994), the methodology used in this study, including the qualitative components, is evaluated in terms of its underlying postpositivist world-view. Hence, the methodology of this study is evaluated in terms of the central and complementary issues of reliability and validity.

Reliability has been defined and subdivided in different ways by researchers from different research traditions, but as defined here follows Neuman (1994, p. 178) in referring to dependency and consistency. Reliability is a necessary but not sufficient condition for validity (Neuman, 1994, p. 186). It has been suggested by Neuman (1994, p. 180) that reliability of quantitative aspects of research can be improved by using clearly conceptualised constructs, precise levels of measurement, multiple indicators, pilot studies and replication. For qualitative research, reliability is enhanced by recording observations consistently, and using multiple data sources (Neuman, 1994, p. 184-185).

Like reliability, validity has been defined and operationalised in a variety of ways by researchers from different research traditions. Here it is taken to mean the accuracy or "truth" of the research findings (internal validity) and whether these findings can be transferred or generalised beyond the particular study (external validity) (Wiersma, 1991, pp. 4-5). In addition, internal validity is frequently subdivided into a number of other types, with those most relevant to this study being criterion validity, within–construct validity, and between–construct validity.

Criterion validity is assessed by comparing an indicator with another pre-existing, "valid" indicator of the construct (Neuman, 1994, p. 183). Within–construct validity is the extent to which different measures of the same construct agree, and between–construct validity is tested by examining the association between two measures of theoretically related constructs (Watkins, 1998, p. 131). Finally, validity has usefully been described by Kvale (1996, pp. 242-244) as a process of checking using a critical approach, questioning to expose different interpretations, and theorising about the appropriateness of methods to research questions.

Research design

At a broad scale, the overall design of this study incorporates a number of measures to maximise reliability and validity. Reliability is enhanced by the triangulated approach to gathering information using complementary methods of interviews and questionnaires, and using clearly conceptualised instruments that have been widely accepted and used in research into higher education (see Chapter 2). In addition, the replicability (external reliability) of this study is enhanced by a number of features advocated by Miles and Huberman (1994, p. 278), such as precise and detailed

description of the context and methods of the study, explicit statement of the underlying worldview of the researcher, and provision of condensed qualitative data to substantiate the conclusions drawn.

The internal validity of this research study has also been enhanced by some of these features of the research design, particularly the use of multiple complementary methods and conceptually connected aspects. Using interviews together with the MSPQ and written responses to investigate related phenomena allows for cross-checking of the validity of both methodological approaches. As stated by Neuman (1994, p. 185) research gains validity when diverse data are used, and when the connections among them are considered. In this study the connections among different qualitative and quantitative data within the two themes of learning approaches and outcome are strongly emphasised, and explicitly sought. The connections between these themes in fact form a third, synthesising study theme.

The major feature of this study that enhances external validity is that it was designed around the activities occurring in a natural setting of learning in first–year biology. Students reported the learning approaches they adopted in a particular topic, and actual assessment items, as well as verbal explanations, were used as measures of learning outcomes. Finally, detailed descriptions of the study context also enhanced external validity by facilitating comparisons with other samples as advocated by Miles and Huberman (1994, p. 279).

Despite these features of the study design, some threats to reliability and validity remain. One particularly intractable threat to the external validity of this study is non-response bias — a problem common to all studies based on voluntary participation. There is inherent tension in this study between the ethical requirements for voluntary participation, and the random sampling requirements of quantitative, nomothetic research. As stated by Burns (1997, p. 18) volunteers differ from non-volunteers, and this reduces the generalisability of the results to the broader population.

Non-response bias could not ethically be prevented in this study where participation was entirely voluntary. Steps were taken to reduce the problem and ameliorate its effects in all stages of the study, ranging from the research questions through to

reporting of results. Firstly, this study has not attempted to establish norms for first–year biology students. The research has been conceptualised as an exploration of learning approaches and outcomes in a specific context and the research questions framed accordingly. Efforts were made to maximise the response rates by explaining the study as clearly as possible and highlighting its potential benefits to other students. Support from the lecturers was solicited and they communicated this to students at the beginning of the study. The students were allowed as much time as possible to fill in the research instruments.

The fact that the study was designed over two consecutive years also potentially affects the validity and reliability of the findings. While collecting data over a single year would have reduced the variability of contexts, it would have had a number of concomitant disadvantages. The major disadvantage would have been a halving of the total sample size, and therefore reduced reliability and validity of the quantitative aspects of the study. Achieving a large enough sample size was an issue in this study, particularly given the uneven number of internal and external students, the spread of data collection points over the course of the unit and the likelihood of incomplete data sets due to attrition or non-response to some aspects of the study.

In addition to increasing the sample size, the two–year design allowed for refinement of ideas and interview questions based on student responses in the first year, and as analysis of the first year's data proceeded. This evolutionary aspect of the study would not have been possible with a single–year snapshot approach. The risk of the two years of data collection was that the two different cohorts may have been different in terms of the variables of interest in this study, that is, in learning approaches adopted or outcomes achieved. This possibility was therefore examined closely in the data exploration phase of the project, the results of which informed subsequent analyses.

The study design is also susceptible in one sense to threats of maturation effects on internal validity (Cohen & Manion, 1994, p. 170). Many events and issues may have impacted on the students and influenced their responses at any point between the two weeks of the topic, the measurement of the approach adopted during that topic, and the measurement of learning outcome from that topic in practical or examination results. For example, in the research design, the item used to measure learning

outcome in the examinations was to be administered a few months after the topic was taught, creating a longitudinal aspect to the study that may be particularly subject to intervening maturation effects. It was partly in response to these potential maturation effects before the examinations that an item to measure outcomes was incorporated into the practical test, which was an integral part of the learning of the topic for internal students, and preceded the administration of the MSPQ. Ultimately, though, because this study is set within an authentic learning context, the period of time between students learning a topic and their later responses to questions about that topic is simply another aspect of what happens in the natural setting of the study. This aspect of the context, like many others, needs to be considered when interpreting the study findings.

In summary, this study design aimed at a compromise between the competing demands of reliability and external and internal validity, recognising that perfect reliability and validity are impossible to achieve (Neuman, 1994, p. 178). External validity was enhanced by using an authentic learning context and outcome measures. Internal validity was enhanced by focusing on single small scale context, which was also the most appropriate and practical scale at which to investigate the study questions.

Interviews

The issue of reliability in interviews within postpositive research is complex. The interpersonal nature of interviews means that they cannot be systematically controlled by the interviewer, because of unavoidable differences in factors such as trust, openness, transparency of communication and avoidance tactics if questioning is too deep (Cohen & Manion, 1994, p. 275). This has led some researchers (e.g., Kitwood, 1977, cited in Cohen & Manion, 1994, p. 282) to suggest that the notions of reliability (and validity) are redundant to interviews, a point shared by many researchers from alternative paradigms. Moreover, if some notion of consistency and dependency of interview results is accepted, measures to improve reliability are likely to threaten validity:

> In proportion to the extent to which 'reliability' is enhanced by rationalization, 'validity' would decrease...In other words the distinctively human element of the interview is necessary to its 'validity'. (Kitwood, 1977, quoted in Cohen & Manion, 1994, p. 282)

The interview procedures used in this study were designed as a "judicious compromise" (Kitwood, 1977, quoted in Cohen & Manion, 1994, p. 282) between the opposing commitments to reliability and validity. The semi-structured interview protocols established a degree of consistency by ensuring that no themes were forgotten by the researcher, and by establishing a roughly consistent order of questions. However, some interviews covered quite a lot of ground that was not in the protocol, or went in unexpected directions.

Reliability of recording interview data was enhanced by audiotaping the interviews to reduce any inconsistency in recording of responses. Although audiotaping interviews may have compromised validity by being threatening to participants, and perhaps exacerbating any response effect, this was considered to be little worse than note-taking (the alternative form of recording responses) and balanced by the advantages of enhanced reliability. Reliability during the analysis of interviews was enhanced by inter-judge discussion and agreement as the coding scheme emerged from the conceptually clustered matrix of transcript data.

Selection bias was a potential threat to the validity of the interviews, particularly given that a relatively small number of students could be selected for this aspect of the study. This issue was addressed by sampling purposively using theoretical criteria. Such purposive sampling, which includes deviant cases, reduces bias and enhances the credibility and potential generalisability of interview studies (Miles & Huberman, 1994, p. 28; Silverman, 2001, p. 253). Another potential threat to validity in interviews is the possibility of response effect (Wiersma, 1991, p. 194), or the tendency of participants to "overstate or understate the 'true value' of an attribute" (Cohen & Manion, 1994, p. 281).

In addition, researchers can inadvertently influence the responses of participants, and focus inappropriately on responses supporting their expectations (Cohen & Manion, 1994, p. 282). A related limitation to interpreting students' explanations of scientific concepts has been raised from a constructivist point of view (Duit, Treagust, & Mansfield, 1996, p. 18). This is that the researcher's constructed interpretations of the conceptions expressed by students may be seriously limited, especially if the researcher is unfamiliar with the range of likely students' conceptions. In this case,

though, the researcher had many years' experience of students' conceptions of meiosis from discussions during practical sessions and tutorials.

The issue of interview interpretation was also addressed by comparing interview data with another measure of the same attribute, as a "convergent validity" test. The triangulated design of the study, therefore, enables the validity of interviews to be measured by the extent of convergence with other measures such as learning approach scores and SOLO levels.

In a discussion of validity in qualitative research, Kvale (1996, p. 241) introduced the notion of that validity depends on the quality of the craftsmanship in the research. This study aimed to maximise the validity of the interviews by aiming for the highest quality craftsmanship. Many of the questions for individual interviews in the second year of the study were piloted in the focus group interviews, which were also used to gauge student (and researcher) reactions to an interview situation, and to particular types of questions. Establishing rapport before and during the interviews was a matter of highest priority. This was partly for ethical reasons (see below) and partly to maximise the information, utility and validity of this aspect of the study. This element of "humanness" is essential for the sort of easy conversation necessary for the validity of interviews (Kitwood, 1977, cited in Cohen & Manion, 1994, p. 282)

In summary, the interviews in this study can be evaluated in terms of reliability and validity, and these issues were considered in planning and conducting the interviews. Semi-structured protocols, audiotaping, purposive sampling, triangulation and attention to interview craftsmanship were all adopted to maximise the reliability and validity of the data gained from student interviews.

Learning approaches questionnaire

This subsection evaluates the reliability and validity of the quantitative aspects of investigating students' approaches to learning. The reliability and validity of the MSPQ instrument are discussed, followed by a critical evaluation of the ways that students might respond to the MSPQ in the context in which it was administered.

Reliability of Learning Approaches Questionnaire

The constructs of deep and surface approaches to learning used in the questionnaire had been conceptualised over several decades of research, and the instrument was

developed, piloted and modified over a number of versions using Australian tertiary student participants (e.g., Biggs, 1987; Prosser et al., 1994). As outlined in Chapter 2, there is widespread consensus about the deep and surface constructs of approaches to learning, and the general reliability of the deep and surface constructs of the SPQ. Nonetheless, some concerns have been expressed about the composition of the surface scale of the SPQ (Christensen, Massey, & Isaacs, 1991; Richardson, 1994b), including the conceptual blurring of different kinds of extrinsic motivation (Biggs, 1987, p. 30; Kember, Wong, & Leung, 1999). These issues are also relevant to the MSPQ, (a close derivative of the SPQ), and point to a possible weakness in the reliability of the instrument used in this study.

The reliability of the four subscales of the MSPQ had been previously tested by administering it to 181 students studying topics of mechanics and photosynthesis in first–year physics and biology at an Australian sandstone university (Prosser et al., 1994, p. 307). These learning contexts had been judged as encouraging deep approaches to learning. The structure of the MSPQ and its subscale reliabilities as established by Prosser et al. (1994, p. 308) are provided in Table 3.7, together with comparative reliability data for the relevant subscales of Biggs' (1987) original SPQ.

Table 3.7: Scale structure and reliabilities of the two–factor version of the MSPQ used in this study, and Biggs' (1987) SPQ

	Cronbach alpha reliabilities			
Instrument and study	Surface Motive (SM)	Surface Strategy (SS)	Deep Motive (DM)	Deep Strategy (DS)
Topic–level 2–factor MSPQ (Prosser et al., 1994), N = 181 [1]	α =.60	α =.61	α =.80	α =.77
3 factor SPQ used to measure learning orientation rather than approach in Australian universities (Biggs, 1987), N = 853 [2]	α =.61	α =.66	α =.65	α =.75

Notes: [1] Modified from Prosser et al., (1994, p. 308), [2] Modified from Biggs (1987, p. 28).

The reliability estimates shown in Table 3.7 accord with results of a number of cross-cultural studies of the original SPQ (in measures of learning orientation) reviewed by Watkins (1998), and are all above 0.50, which is considered "acceptable for a research instrument used for group comparisons" (Watkins, 1998, p. 131). As shown in Table 3.7, the two surface scales show the lowest reliability, which is consistent with literature norms. It was decided not to pilot the MSPQ in this study as it had been shown to be reliable by other researchers in such a similar context.

Construct validity of the MSPQ

Methods for evaluating the validity of learning approaches questionnaires have been discussed by Watkins (1998, p. 131). He argued that demonstrating the construct validity of an instrument requires within–construct validity, commonly assessed by factor analysis of scales, and between–construct validity, which is assessed by a range of measures of relationship between learning approach questionnaire scales and other variables.

The majority of studies have strongly supported the within–construct validity of the deep and surface constructs of the original SPQ, and by extension, the MSPQ which is so closely derived from the SPQ. The robustness of the surface scale, though, has been questioned (e.g., Christensen et al., 1991; Richardson, 2000, p. 84). Richardson (1994a, p. 462) stated that studies involving the SPQ that he had reviewed did measure a "generalised surface approach and a generalised deep approach to studying". He suggested also that an identical structure had been found in Nepal, Nigeria and the United Kingdom and that the deep and surface approaches had some cross–cultural validity.

In a review of the within–construct validity of the SPQ, Watkins (1998, pp. 131-136) summarised factor loadings based on two–factor solutions to principal axis factor analyses of samples from eight countries. He concluded that, apart from the achieving approach, the results were "clear-cut with distinct surface and deep approach factors" (Watkins, 1998, p. 131), but raised some questions about its applicability to Nepalese and Chinese students. Kember and Leung (1998) tested the dimensionality of the SPQ on a sample of over 4000 university students in Hong Kong, and concluded that a two–factor model differentiating deep and surface learning approaches was the most appropriate. A two–factor version of the SPQ has been developed (Biggs, Kember, & Leung, 2001), maintaining the deep and surface scales, each comprising motive and strategy subscales.

Although the studies cited above support the within–construct validity of the deep and surface approaches of the SPQ, most of these studies used the original three–factor version of the SPQ, administered in the context of learning orientation rather than learning approach. Richardson (2004, p. 355) also pointed out that the content

validity of instruments such as the MSPQ is questionable because of massification of the higher education system since their development in the 1970s.

None of the published studies using the MSPQ have reported results of factor analysis of the scales (e.g., Crawford et al., 1998a, 1998b; Hazel et al., 2002; Prosser et al., 1994; Trigwell et al., 1999). Therefore, although acceptable within–construct validity of the MSPQ is suggested by the robustness of the constructs in its predecessor the SPQ, further information on the construct validity of the MSPQ was sought from its authors. This is shown in Table 3.8.

Table 3.8: Varimax rotation of factor solution for MSPQ items

Scale	Item	Factor 1	Factor 2	Factor 3
Surface	1		.44	
	3		.39	
	5		.51	
	7		.46	
	9		.41	
	11		.40	
	13		.39	
	15		.44	
	17		.50	
	19		.40	
	21		.45	
	23		.54	
	25		.41	
	27		.64	
Deep	2	.64		
	4	.38		.43
	6	.48		
	8			.55
	10	.54		.40
	12	.37		
	14	.73		
	16			.64
	18	.64		
	20	.70		
	22	.57		
	24	.69		
	26	.56		
	28			.70
% variance explained		16.7%	11.3%	8.9%

Notes: Only loadings above 0.3 are displayed

Table 3.8 presents the results of a Principal Component Analysis conducted on the MSPQ, which had been administered in the context of a short topic in first–year biology at an Australian university (Michael Prosser, personal communication, 20 August 2002). As shown in Table 3.8, all the surface approach items loaded on Factor 2, while the deep approach items were split between Factors 1 and 3, and the

three factors explained a total of 37% of the variance. The variance explained by these three factors is somewhat less than the 50-60% of variation typically explained by factors in factor analysis (Meyer & Shanahan, 2003, p. 9), but the factor structure provides reasonable support for the validity of the deep and surface constructs of the MSPQ.

The between–construct validity of the MSPQ has been clearly demonstrated. In the topic–level study in first–year science using the MSPQ (Prosser et al., 1994), factor analysis of the relationship between scales on the MSPQ and a modified student experience questionnaire was consistent with expectations. The results indicated that deep approaches associated with perceptions of good teaching, clear goals and independent learning, while surface approaches associated with perceptions of unclear goals and inappropriate workload (Prosser et al., 1994). A study into learning approaches in first–year mathematics (Crawford et al., 1998a) used a version of the MSPQ slightly reworded to fit the context (i.e., items included the word "mathematics" instead of "this topic"). Again, between–construct validity was apparent in results of factor analysis of the MSPQ scales with a number of other variables. Deep and surface scales were associated with the other variables as had been predicted (Crawford et al., 1998a, p. 463). Factor analysis of the results of a study of nearly 4000 science students in Australian universities (Trigwell et al., 1999) also supported between–construct validity, indicating strong relationships between the constructs of the MSPQ and teachers' approaches to teaching.

Validity of information from learning approaches questionnaire

The validity of information from questionnaires such as the MSPQ depends not only on the validity of the constructs within the instrument, but also on the way participants respond to the instrument. Threats to the internal validity of information gained from questionnaires such as the MSPQ come from a variety of sources.

One of these threats is the problem of "response sets", that is, students responding in a particular way to all items, for example, always tending to "agree" rather than "disagree". This relates to the problem discussed by Krosnick (1999, pp. 546-552), that answering questionnaire items optimally involves considerable cognitive effort, and some respondents become fatigued or unmotivated and tend to respond superficially to items. As pointed out by Richardson (2004, p. 354), the SPQ does not

control for response bias, as all items have the same polarity (i.e., there is no reverse scoring); however, it has been claimed by Waugh and Addison (1998, p. 97) that positively worded items are preferable for modern measurement models such as Rasch. The MSPQ consists of four subscales with theoretically opposite and relatively obscure "positive" directions. Therefore response sets, if they occurred, would result in mixed surface and deep scores rather than extremes of one approach. The tendency to agree with propositions is less likely when questionnaires are anonymous (Wiersma, 1991, p. 285), however, anonymity was not possible in this study because questionnaire responses needed to be compared with learning outcomes measures.

An additional potential threat to the validity of the MSPQ considered in this study relates to its use with external students. The development of the SPQ and most subsequent research has been in the context of on–campus students, and Richardson (2000, p. 175) has cautioned against the "unreflective" use of instruments designed for on–campus students in research into off–campus students. For example, a study into distance learning students in Hong Kong/Macau by Ekins (1992) using the SPQ attracted criticism from Richardson (2000, p. 80) for the following irrelevant item: "I learn best from teacher(s) who work from carefully prepared notes and outline major points neatly on the blackboard.". A slightly modified version of this item also appears in the MSPQ used in this study (item 19). Because, however, the externally enrolled students in this study attend compulsory face-to-face sessions in residential schools, and are provided with notes written by their lecturers, this item seems equally relevant for internally and externally enrolled students. In this and other items of the MSPQ (see Appendix F), there are no references that seem to be inappropriate to either group of students.

The use of the MSPQ across both cohorts is supported indirectly by the results of a study of over 2,000 students at the Open University (Richardson et al., 1999). The study found that distance students' responses to the ASI were equivalent to findings for on–campus students. This does suggest that the deep and surface constructs, as articulated in the ASI, are equally valid for distance students. It still may be the case that the MSPQ may be differentially appropriate for internal and external students, but as stated by Richardson (2000, p. 175), it was necessary that the same instrument be used for all students.

A threat to the validity of the quantitative data on learning approaches is the issue of the focus of student awareness when responding to the MSPQ, particularly given its use across internally and externally enrolled students studying the same unit. In a sense these two different groups of students are experiencing quite different learning situations, and responded to the MSPQ at different points in their study of the topic. Students' study of a topic takes place in a variety of places and times, for example, in the practical sessions, at home and before related assessment. For the external students, in particular, the home study context is very different to the residential school context, where the MSPQ was administered. The learning approaches of any individual student may well vary across these study sub-contexts and their responses to the MSPQ may therefore depend to some extent on which aspect of their study they have in the forefront of their mind. In addition, the fact that the MSPQ wording uses the present tense may also have lead some internally enrolled students (who had recently studied the topic for a practical test) to respond in the context of the next topic in the unit, despite clear oral and written instructions to the contrary. Changing items to the past tense, though, would have been inappropriate for external students.

A further related issue is that the MSPQ is subject to known limitations of respondent's self-reports of their behaviours and attitudes. Studies reviewed by Richardson (2004, pp. 354-355), on questionnaires such the ASI and SPQ suggest that "under some circumstances students reconstruct their autobiographical memories" and these reconstructions may be what they report on the questionnaires. This is a problem especially when these reports are solicited retrospectively.

The issues described above are not unique to this study, and represent known weaknesses of quantitative methods of collecting information about human behaviours and attitudes. A number of features of the study design, though, acted to counter or detect these potential threats to internal validity. The fact that a single, short–term topic was chosen, helped to limit the focus of the students' responses. In addition, this focus was transmitted to students by explicit print and verbal instructions to respond only in the context of the target topic, and to respond taking into account everything that they did in relation to that topic. The validity questionnaire was incorporated specifically to check that the students remembered the learning context, and to identify if they approached their study any differently in the subsequent topic.

The major aspect of the study design that ameliorates the weaknesses of the quantitative methods described above is the inclusion of more qualitative methods for eliciting information about the learning processes of the students in the study. The interviews with students allow for deeper exploration of the students' study behaviours and motives, checking of their focus and prompting if necessary to ensure the discussions focused on all aspects of their study in the target topic. Reporting of interviews with "thick" description and low–inference descriptors (LeCompte & Preissle, 1993, p. 338), and cross–validating findings of quantitative and qualitative aspects of the study are features of this triangulated research design. These combine to maximise the reliability, validity and utility of the findings relating to students approaches to their learning.

In summary, the reliability and validity of the SPQ and its topic–level derivative used in this study have been supported in numerous studies into student learning. Threats to the validity of data collected from the MSPQ in this study relate to response sets, potentially different validity for internal and external students, the focus of awareness in students' minds when responding to the items, and known limitations of self-reporting instruments. These issues were addressed as much as possible, by choosing a single topic as the study context, using a validity check, and by triangulating MSPQ responses with interview data.

Measures of learning outcome

This subsection evaluates the reliability and validity of the methods used to determine students' learning outcomes. The reliability and validity of the SOLO model are discussed, and the more general methodological issues associated with the learning outcomes component of this study are evaluated.

Reliability of SOLO model

Like the learning approaches questionnaire, the constructs used in the SOLO taxonomy have been clearly articulated in a number of studies reviewed in Chapter 2. Categorisation of responses in SOLO is guided by criteria of the abstractness of response, as well as the structural complexity of responses. As argued by Biggs and Collis (1982, p. 186), inter-judge agreement in this process is a crucial determinant of reliability for SOLO, where consistency of responses across items and over time is theoretically improbable.

Reliability data for the original single-learning-cycle of SOLO have been reported in a number of studies. Biggs (1982, p. 188) examined inter-judge agreement for three different disciplinary areas and concluded that "the results to date are however quite acceptable, and indicate that, whatever SOLO is measuring, different judges agree about the measurement.". In a study of first–year university psychology students (Van Rossum & Schenk, 1984), inter-judge agreement of 91% was established for SOLO ratings of 69 pairs of written responses to open–ended questions. Inter-judge agreement of 78% was established for SOLO levels in a study of 14 postgraduate education essays (Boulton-Lewis, 1992, p. 486). One study has reported problems with inter-judge disagreement in applying SOLO to postgraduate essays (Chan, Tsui, & Chan, 2002, p. 517), and these researchers hence questioned the reliability of SOLO, suggesting that the categorisation criteria were conceptually ambiguous and led to unstable categories. Most other studies using earlier versions of SOLO reported in Chapter 2 have not questioned the reliability of the model.

In terms of the two-learning-cycles version of SOLO, Panizzon (1999, p. 328), has reported intrarater and interrater reliabilities of 98% and 97% respectively using 20 random samples coded against SOLO level descriptors. This was in a very similar context to the present study, that is, a study into the understanding of biological concepts of first–year university and senior secondary students. By contrast, difficulties were experienced in applying the two-learning-cycles version of SOLO (Redden, 1995, p. 282) to children's arithmetic responses.

In this study, the procedure used to categorise responses using SOLO incorporated three judges throughout the process, as well as at the end. This paralleled a method advocated by Bowden (1992 [1994], cited in Sandberg, 1997, p. 206), to improve inter-judge reliability in establishing categories of description of individuals' conceptions in phenomenography. According to Sandberg (1997, p. 207), Bowden argued that in defending the developing categorisation to co-researchers familiar with the data, "different interpretations can be exposed until intersubjective agreement is achieved". The procedure used in this study did expose different interpretations. The combination of perspectives, imposed by different levels of familiarity with the subject matter and with SOLO, was an advantage in ultimately establishing a stable pattern of responses, and agreement on the fit of the SOLO

model with the data. Reliability was also investigated using an inter-judge test at the end of the coding process.

Validity of the SOLO model

The criterion validity of the original version of SOLO was initially examined (Biggs & Collis, 1982, pp. 189-192) using two measures. One of these was a test of the correlation between SOLO ratings of 100 Grade 9 essays and independent ratings of the work by an experienced teacher, which found that SOLO ratings did define what the teacher meant by quality. In addition, factor analysis of intercorrelations between SOLO levels and other tests suggested that this version of SOLO was also related to other "cognitively relevant processes" (Biggs & Collis, 1982, p. 192). Strong evidence for the criterion validity of SOLO is also apparent in the coincidental correlation between SOLO levels and Marton's four hierarchically related levels of learning (Biggs & Collis, 1982, p. 14). The validity of the original version of SOLO has also been demonstrated by the correlation between SOLO levels and phenomenographic analyses in the context of first–year science students understanding of photosynthesis (Hazel et al., 2002, p. 748).

In addressing the between–construct validity of the original version of SOLO, Biggs and Collis (1982, pp. 192-199) tested the association between SOLO, achievement and personological factors related to learning orientations for 329 Grade 9 students. They concluded that SOLO was related to school achievement, and that high SOLO levels were associated with students demonstrating a deep learning orientation. This result supported earlier work that had examined the association between SOLO levels and study processes (the conceptual forerunner of Biggs' learning "approaches") for 69 university undergraduates. This study found that high SOLO levels were obtained by students adopting a meaning (deep) orientation (Biggs, 1979, p. 390). Since then, the between–construct validity of the original single-learning-cycle version of SOLO has been supported by a number of studies which have provided interpretable and theoretically coherent relationships between SOLO levels and learning approaches or orientations in university students (e.g., Hazel et al., 2002, p. 745; Tang, 1994; Trigwell & Prosser, 1991, p. 272; Van Rossum & Schenk, 1984).

Although there is an accumulation of evidence indicating the validity of the single-learning-cycle of SOLO, the validity of the more recent two-learning-cycles version of SOLO has been less well reported. To date, no study has examined the relationship between SOLO levels obtained from the new model and theoretically related concepts such as learning approaches to test between–construct validity.

Validity of information on students' learning outcomes

This study was designed in part to examine the variation in students' learning outcomes according to the most recent version of the SOLO model, and to examine the relationship of this to learning approaches in the real–life natural setting of a first–year biology unit. This has some consequences on the method chosen to assess students' learning outcomes.

Perhaps the most important of these is the use of a single examination question to indicate SOLO categories of learning outcomes of internal and external students, with the addition of a practical test question for internal students. The SOLO model is a theoretically based framework for qualitative analysis of responses to particular questions. Qualitative analysis using SOLO (or other perspectives such as phenomenography) is a time–consuming and complex process, with a single question which is worded to elicit extended responses being an appropriate scale for analysis. This approach has precedents in many comparable previous studies (e.g., Hazel et al., 2002).

The choice of using an actual standard unit assessment item as a measure of outcome will have influenced the outcomes detected in this study. Students' responses to a genuine assessment question are likely to be different to their responses to the same question administered as an item of an extracurricular research instrument. Although responses in the genuine context may well reflect the influence of assessment on approaches to learning and therefore outcomes, this option is a more valid indicator of the outcomes of students' actual learning activities.

In summary, this subsection has evaluated the methodology used in investigating students' learning outcomes. The reliability of the SOLO model has been supported in a range of previous studies by inter-judge tests, which were also conducted in the current study. The validity of SOLO is supported by its coincidental correlation with

Marton's levels of learning outcome, and with subsequent phenomenographic analyses of learning outcome. These studies, though, relate only to previous versions of SOLO. The reliability and validity of the learning outcomes aspect of this study have been enhanced in some respects by use of actual assessment questions, although this has limited the number of questions that could be asked.

Ethical considerations

This subsection critically evaluates some ethical issues inherent in the research design. All aspects of this study were presented to and approved by the university Human Research Ethics Committee (HREC), approval number HE01/150. Therefore in the judgement of the committee, the study complied with standard ethical considerations such as voluntary participation, fully informed consent, the "do-no-harm" principle and confidentiality.

Nonetheless, some potential ethical issues surrounding this research methodology warrant further exploration. For example, students who were interviewed in this study participated in this aspect of the project voluntarily, with informed consent, and with the specific request in individual interviews that they ask for the interview to stop if they found it stressful. However, even with these protections, the interviews could be criticized on ethical grounds. Fontana and Frey (1994, p. 373) argued that "most traditional in-depth interviewing is unethical" because interviewing techniques are actually manipulating respondents. Neuman (1994, p. 120) raised the question of whether it is ethical to cause stress or loss of self-esteem. Both of these were issues which were possible in the types of interviews conducted in this study.

This is a difficult issue given the crucial role interviews have in much educational and other social research. As Fontana and Frey (1994, p. 374) also stated, "a question must be asked person-to-person if we want it to be answered fully", and as argued by Neuman (1994, p. 119) there is a need to balance the value of the research against the value of non-interference in the rights of others. In this case it was judged that the value of interviews outweighed the potential stress they might place on interviewees.

The researcher, though, was extremely careful to develop real rapport with the students, to minimise pressure, not to probe uncomfortably deeply or persistently in knowledge–based questions, and to respond to students in a way most appropriate to

them. For instance, if a student was looking anxious and hesitating over a question they may have been told that they did not have to answer it if they did not want to, and so forth. The researcher also told the students at the beginning of the interview that she would be happy to answer any questions about meiosis afterwards, and many participants took advantage of this to ask questions to clarify their understanding. As advocated by Kvale (1996, p. 155) this was incorporated as "fair return" by the researcher for the students' participation.

A related ethical concern was that, on the advice of the HREC, the information sheet directed students to the university's Academic Skills Office (ASO) if they became stressed about their studies or progress while participating in the research. This provision, though, was ethically complicated by the fact that the researcher herself was employed in the ASO. However, it was known to these students that their first port-of-call for advice on study issues was another ASO staff member who was based in the science faculty with prime responsibility for first–year students. It was therefore considered unlikely that this complication would hinder students accessing help if required.

One particular ethical issue in this study was the examination and practical test items designed to gauge students learning outcomes from the topic, and which related to the non-participants as well as participants. While including more outcome questions may have been theoretically desirable on some grounds, it would have been ethically unjustifiable to depart from standard practice in the unit, especially considering that non-participating students were sitting the examinations and tests.

Finally, the reporting of this research study, like many others, poses some challenges to the ethical issue of confidentiality. Lecturers as well as students were participants in the study, and it would be theoretically possible for some readers to identify the lecturers involved if they were so inclined. It was not possible to report the research accurately and clearly without revealing the identity of the institution and academic area involved in this study. The researcher was at times uncomfortably aware that judgements could be made on the quality of the institution's teaching or its student body on the basis of this thesis. This remained an unresolved ethical difficulty given that the whole institution could be described as a *de facto* participant, by virtue of the

participation of some of its members, but without the benefits of confidentiality accorded to direct participants.

In summary, some ethical concerns are difficult to avoid in conducting research into human attitudes and behaviours, and steps were taken to minimise any ethical problems. In the main these were assuring fully voluntary, informed and confidential participation, and limiting examination questions to the standard number and type. The research was conducted in a spirit of ethical responsibility, and awareness of its potential impact on participating and non-participating students, and the institution.

Summary

This section has taken a critical approach to evaluating the methodology used in the research study. The issues of reliability and validity were considered in evaluating the overall research design, as well as the more specific methodologies associated with interviews, MSPQ responses and students' responses to meiosis questions. Finally, the ethical implications and limitations of the study were discussed. The triangulated research design, use of strongly supported and theoretically well–resolved models and instruments, and attention to craftsmanship and ethics were all adopted to maximise the reliability, validity and ethical integrity of this study into student learning.

Overview

The discussion of the research methodology outlined in this chapter focused initially on establishing the context of the research, as an essential background to the students' learning approaches and outcomes. Included in this discussion were aspects of the unit and topic organisation and content, as well as an explanation of meiosis, which is the conceptual context for investigating students' learning outcomes. The description of research design then outlined data collection procedures, including issues relating to recruitment of participants and the details of the study sample and response rates. The research instruments used in the study were discussed in terms of the theoretical justification for their use and a description of their implementation in the study context. The analytic techniques used in this study span a range of qualitative and quantitative techniques. Theoretical issues relating to these analyses were described, followed by specifics of data analysis for each research question.

Finally, the study methodology was critically evaluated, including aspects of research design, data collection instruments and ethical issues.

The methods outlined in this chapter yielded quantitative and qualitative results relating to the three research themes of students learning approaches, outcomes, and the relationship between these two components of the 3P model. These results are presented in the following three chapters. Chapter 4 outlines results pertaining to learning approaches, Chapter 5 describes learning outcomes, and Chapter 6 presents results elucidating the relationship between learning approaches and outcomes.