Chapter 4
Research Method

> *We do no violence to objectivity by knowing what we want to look at, by locating ourselves in a setting where we are most likely to see it, by being well acquainted with the work and procedures of others whose methods or focus are of possible relevance to our proposed research, by having theoretical/conceptual frameworks within which to couch our observations, or by having hunches about how we think a system works* (Wolcott, 1988:25).

Overview

The investigation of mitigation followed a three-pronged approach. The initial and primary avenue of investigation was linguistic, the aim being to generate an empirically-driven typology of supervisory mitigation. The second prong used an ethnographic approach, seeking to understand the feedback process, and especially the face-threatening act of criticising, from the perspective of the supervisor so as to establish a data-driven description of supervisory concerns about feedback. The third prong was experimental, and here attention shifted - in terms of perspective (from the supervisor to the student teacher) and in terms of method (from qualitative to quantitative) - with the aim being to investigate perceptions of mitigation in simulated supervisory conferences, under controlled conditions. Chapter 4 is devoted to the research method of the first and second prongs; the third prong is outlined in Ch. 7.

4.1 Introduction

4.1.1 A three-pronged approach: triangulation

4.1.1.1 Concurrent processes

The three aspects of the research were three prongs rather than three phases. It would be wrong to consider it a neatly divided, step-by-step, linear process with three distinct temporal stages. Although the initial step was linguistic, and the next to be begun ethnographic, and the

last started and completed experimental, the three prongs in fact continued for a large part simultaneously, with data collection for all three extending over equivalent time span. In this way the process echoes the simultaneity of data collection and analysis typical of qualitative research (Merriam, 1991).

4.1.1.2 Between-method triangulation

The three prongs provide internal validity to the investigation through triangulation (Denzin, 1970; Webb, Campbell, Schwartz & Sechrest, 1965), a research protocol through which the same phenomenon is investigated by different methods, different data sources, or by different investigators. Investigator triangulation was largely impossible because of the nature of the project as doctoral research; however, other researchers were at times solicited e.g. an inter-rater check in the application stage of the typology; and member checks and phenomenon recognition in the ethnography. Triangulation of data was achieved through a range of data sets: 10 supervisors in the supervisory dialogues in the linguistic phase (Ch. 5); another 30 in the ethnographic interviews and a further 40 in the grounded survey (Ch. 6); and 231 subjects in the controlled experiment (Ch. 7).

However, the major triangulation undertaken is the between-method operation (Denzin, 1970), combining three dissimilar methods of collecting data. The audio recordings in the linguistic phase yielded verbatim transcripts, processed through a grounded theory approach. The interviews in the ethnographic study yielded a corpus of 'data sets' (Baxter, Eyles & Willms, 1992:215) which was content analysed. The questionnaires in the controlled experiment yielded largely quantitative data, processed by statistical aggregation; and some qualitative data which was content analysed.

4.1.1.3 Proximity to the experience of supervision

A valuable way of perceiving the three prongs is in terms of the degree of closeness to the experience of supervision. The transcripts of the dialogues are the primary source, for this is as close as one can get to the experience itself (Fairclough, 1989; Gilbert, 1992). However, as discourse is itself an interpretation of experience, the transcripts in fact serve as a record of 'a pre-interpreted domain' (J.Thompson, 1984:133). The dialogues are 'real' in that they occurred naturalistically, as part of normal procedures in teacher education courses and language teaching contexts; and were not set up for the purposes of the study (Grimshaw,

1974) or in any way 'introspected' (Burns, 1992/3:61). Rather, a tape-recorder was switched on to record a process as it would normally happen in institutions where supervisory processes are 'part of the territory'.

One step further removed from the core experience (Gilbert, 1992) is the ethnographic investigation which attempts to 'get inside the head' of supervisors in order to find out what motivates and concerns them. This method is both introspective in that it probes invisible processes - thought-processes, attitudes, decision-making - and retrospective, in that it calls on subjects' recollection of past experiences of supervision events. The third prong, the experimental study, is yet a further step removed, for here, through simulation, the attempt is made to discover respondents' perceptions of simulated supervisory language. The experiment is conceived within the hypothesis-verifying paradigm, using a controlled situation and manipulated events; and offers a limited, albeit readily quantifiable, perspective. Part of its value is that it allows one, through the triangulation, to blend quantitative and qualitative methods, and so reduces the 'dependence on talk' as a means of 'making sense of the setting' (Miles & Huberman, 1991:234).

4.1.2 Why triangulate?

Triangulation is a logical procedure in the direction of ensuring validity in the absence of any external measure against which to measure findings: one looks therefore for corroborating evidence or convergent findings in other internal indices (Miles & Huberman, 1991). The advantages of a multi-method approach in educational research warrant review here .

Firstly, the complexity of human experience and of the contexts in which human beings interact and make their meanings manifest, is such that the single method (such as the hypothesis-verifying observation characteristic of fields like medicine and chemistry) can offer only limited insights. At the same time, the very flexibility and outward-seeking, non-verification orientation of qualitative research attracts the charge that it is loose, anecdotal, impressionistic, subjective, biased and 'soft'. The 'convergent validity' of a multi-method approach may be 'a form of robustness' (Fielding & Fielding, 1986:28). If, as Fielding and Fielding contend, every information-gathering mechanism is both 'privileged' and 'constrained' by features of its own nature (1986:31), then a blend of different mechanisms is a source of enrichment.

An additional factor is the impact a research method has on the investigator. Methods are not atheoretical or neutral, but rather, operate as mediating 'filters' through which reality is selectively experienced (Cohen & Manion, 1989; Smith, 1975). Similarly, Ratcliffe comments: 'data do not speak for themselves' (1983:149); numbers, equations and words 'are all abstract, symbolic representations of reality, but not reality itself' (1983:150). There is no one 'truth' (Fielding & Fielding, 1986:25). Exclusive reliance on one method of investigation may therefore distort the investigator's view of the 'slice of reality' with which he or she is concerned (Cohen & Manion, 1989:269). Webb et al. (1965) who coined the term 'triangulation', spoke of validating a finding by subjecting it to 'the onslaught of a series of imperfect measures' (cited in Miles & Huberman, 1991:234). This notion of the unique fallibility of every method - of a method-specific, distorted representation of reality - is a frequent refrain in research method literature, perhaps best captured in the line: 'our truth is the intersection of independent lies' (Levins, 1966, cited by Fielding & Fielding, 1986:23). As each method has its own bias, one should, in place of avoiding the issue, actively 'combine methods which have different biases' (Stubbs, 1983a:236). A similar sense of vulnerability can be gleaned from the metaphorical language of Campbell and Fiske's description of triangulation - that which happens when 'a hypothesis can survive a confrontation of a series of complementary methods of testing' (1959, cited in Fielding & Fielding, 1986:23-24).

Thirdly, triangulation through multiple methods allows one to counter the problem of 'method-boundedness', predicted some forty years ago by Boring, who wrote (1953, cited by Cohen & Manion, 1989:270):

> As long as a new construct has only the single operational definition that it received at birth, it is just a construct. When it gets two alternative operational definitions, it is beginning to be validated. When the defining operations, because of proven correlations are many, then it becomes reified.

Triangulation alone does not ensure internal validity to a set of findings, because each procedure must itself have its own validity. The overall chain, then, is no stronger than its weakest link (Pierce, 1936, cited by Fielding & Fielding, 1986:28). Consequently, in the description of the three prongs of this study, attention will be paid in each case to the notion of internal validity within each method.

Perhaps the greatest value offered by the protocol of triangulation is the creation a particular researcher's mind-set. It can transform investigators into their own most exacting critic -

requiring them to test, question, re-appraise, consider alternatives, play devil's advocate, identify Achilles heels and other such classic, detective-like, counter-checking, investigatory strategies (Miles & Huberman, 1991:234). In theory, this process is infinite, what Cicourel terms 'indefinite triangulation' (1973:124). The multifarious checks against threats to validity - within-method as well as cross-method - lessen 'recourse to the assertion of privileged insight' (Fielding & Fielding, 1986:25). Cook and Campbell counsel researchers to 'trenchantly' examine all imaginable threats (1979, cited by LeCompte & Goetz, 1982:50). Miles and Huberman (1991:235) concur:

> Triangulation is a state of mind. If you self-consciously set out to collect and double-check findings... the verification process will largely be built into the data-gathering process, and little more need be done than to report on one's procedures.

4.1.3 The audit trail

The purpose of this chapter is to provide an 'audit trail' (Merriam 1991:172). It is hoped that a clear pathway of procedures used - linguistically, ethnographically and experimentally - will provide independent judges or other researchers with the information by which they are able to evaluate findings. In their discussion of reliability, Lincoln and Guba suggest that qualitative research would be better to offer 'dependability' or 'consistency' (1985:288). Rather than demanding that outsiders obtain the same results, one wishes that outsiders will agree that, given the data, the results make sense. Along with a clear statement of the investigator's position and the protocol of triangulation, the audit trail offered in this chapter is intended to offer research reliability (Goetz & LeCompte, 1984; Lincoln & Guba, 1985; Merriam, 1991).

More attention has deliberately been given to the linguistic procedures relative to the ethnographic and the experimental. In the first place, the three prongs are not of equal import: the linguistic prong provides the primary data source, the other two providing ancillary, corroborating evidence. In the second place, the conceptual framework guiding the approach - drawing on strengths and interests in the researcher's academic background, and generating the research question - is linguistic: the speech event of the SD is equated with its language. The point of departure for the investigation is linguistic - an 'intuition' that the fragility of the SD will reveal itself under linguistic scrutiny. In addition, the major contribution that this study seeks to make is linguistic, providing insight into supervisory communication and applied outcomes connected to supervisory training in communication skills. Another factor is that

because discourse analysis is a relatively new procedure, lacking a well-defined and established methodology, the onus is on the researcher to provide ample guidance and information about both data collection and analysis processes (Sinclair & Coulthard, 1975; Stubbs, 1983a). Lastly, as the critical review has shown, supervision has traditionally been an educational concern and therefore more detailed signposting in the linguistic descriptions may be appreciated. For these various reasons, then, both in the description of method and findings, the linguistic prong is favoured with weightier emphasis.

4.2 The first prong: research method in the linguistic study

> The discourse analyst treats... data as the record (text) of a dynamic process in which language was used as an instrument of communication in a context by a speaker... to express meanings and achieve intentions (discourse). Working from this data, the analyst seeks to describe regularities in the linguistic realisations used by people to communicate those meanings and intentions (Brown & Yule, 1983a:26).

4.2.1 Preliminary analytical issues

4.2.1.1 The critical FTA: the minimal unit of analysis

This section takes up the threads begun in the latter part of Ch. 3. As outlined there, the felix culpa of the pilot study revealed that despite the discoursal power being almost totally in the hands of the supervisor, there was an overwhelming impression of restraint at the linguistic level of the utterance. This was perceived as a pervasive reluctance on the part of the supervisor to express a negative view; an unwillingness to be totally frank; a tendency to mollify and minimise; and because of all this, an inclination towards convoluted and marked usage. The sense was of utterance-level mitigation pulling against or tempering the unleashing of power.

The first step, then, was the recognition of restraint in the discourse and the identification of the rubric of mitigation as its linguistic realisation. The next step was 'combing' the pilot data for FTAs specific to the SD as a particular activity type - namely, directly criticising; making a request for action; and suggesting/advising. These were defined as a sub-set of speech acts, all functionally connected to delivering criticism.

It seemed appropriate to nominate criticising as the minimal unit of analysis. A number of factors played a part here. Firstly, it was clear from the pilot study that such acts provoked intense clusters of mitigation. At such moments, a great deal of effort seemed to be expended to avert threat and promote harmony. This notion - essentially Candlin's concept of 'interactional cruces' (1987:415) - suggested that such acts may be 'sites of contestation and struggle' (Giroux, 1985:33), and worthy of closer scrutiny.

This seemed consonant, too, with the awareness, outlined in Ch. 2, that misgivings about criticism are central to supervision. Central here is the sense that criticism is the *most* face-threatening act within the speech event. From the viewpoint of the teacher, the worst thing that can happen is to have your teaching criticised; and from the viewpoint of the supervisor, the worst thing you may have to do is to criticise the teacher's performance. Narrowing the unit of analysis to the act of criticising, therefore, seemed very pertinent to the activity of supervision.

Concomitantly, the on-going literature search into supervision revealed that the issue of criticism was important precisely because it was problematic: while central to the goal of supervision as instructional improvement, it was often skirted in the practice. Two points emerge here: one was the discovery that criticism was being avoided by supervisors during the speech events of supervision; and the other was that the issue itself had not been closely attended within the research community. In a sense, this was a double skirting, together justifying a focus on the speech act of criticising.

Accompanying this was the researcher's own growing perception (heightened by a reading of critical linguistics e.g. Fairclough, 1989) of the SD as an institutional event replayed repeatedly in institutional settings. This is the sense that the SD does not have the nature of (and therefore ought not be treated as) a one-off, isolated event, in any way unique or idiosyncratic; but rather as a recurring set of patterns of interaction, firmly embedded in institutional mores and conventions, motivated by and serving institutional ends - a generic potential played out repeatedly wherever teacher supervision occurs. The obvious attraction here is to interpret the tension between discoursal power and linguistic restraint within an institutional framework: while the institutionally imposed rules and obligations of the supervisor sanction the delivery of criticism, it is the individual supervisor's 'squeamishness' (experienced through face constraints within the encounter) that accounts for the psychological restraint and the resultant linguistic mitigation. Mitigation, then, is the evidence in the supervisor's language of how they

resolve the clash-of-goals inherent in the implementation of their institutional duties. This means that the two 'frames' (Goffman, 1986) - institutional and individual - must be activated in the assignment of meaning to the social event of supervision. It was considered that determining the minimal unit of analysis as the critical FTA would assist this investigation.

With the decision to adopt the critical FTA as the minimum unit of analysis, the stage was set for the next step. The pilot SD - now renamed [SD1] - was combed for its critical FTAs, operationally defined as utterances bearing a direct or indirect element of negative judgement on the part of the supervisor in regard to any aspect of the lesson observed. Following Brown and Gilman, the length of the FTA was defined as 'all the text necessary to specify the FTA plus all continuous, antecedent and subsequent text that does not belong to the new speech act (1989:174). Ten such FTAs were identified in [SD1]; and a subsequent validity check successfully confirmed that the researcher's perception of what constituted critical FTAs matched the supervisor's [SD1] *post factum* perception.

The validity check match-up perhaps masks the difficulty involved in identifying and classifying FTAs. Rarely are such messages delivered straight-forwardly - openly, frankly, and without qualm-induced 'lumps in the carpet'. Within the register of the SD, there is, of course, no performative verb construction - 'I criticise you for...' - comparable to the performatives used to baptise children, marry couples, absolve sinners, pronounce verdicts, declare wars, bequeath possessions or name ships. Indeed, this thesis is largely a data-driven testimony to equivocation - the intricate avoidance of frankness in criticism. As such, the FTAs are not as readily retrievable as perhaps the giving of praise might be[1]. Unearthing and labelling the critical FTAs was aided by a degree of linguistic sensitivity derived from extensive experience in the role of supervisor. Such experience affords a certain familiarity (albeit a user's familiarity, if not, at the early stages, a researcher's) with the SD *qua* speech event. In addition, what is needed, as the next section outlines, is an awareness of the theoretical and practical issues involved in speech act analysis.

[1] This is not, of course, to suggest that acts of praise are not themselves infused with their own unique complexities (see App. 21).

4.2.1.2 Recognising acts of criticism

4.2.1.2.1 Speech acts

Identifying any speech act for what it is poses not inconsiderable difficulties, as critics of speech act theory have pointed out (Conte, 1981; Franck, 1981; Levinson, 1979; Levinson, 1981; Rosaldo, 1982; Streeck, 1980; Stubbs, 1983a). A primary difficulty derives from the complexity of analysing the 'raw data' (Coulthard, 1977:9) of naturally occurring discourse, as distinct from the idealised, decontextualised, 'self-contained', non-interactive language on which philosophers of speech acts have mostly based their work (Coulthard, 1977:9).

In confronting real language used by real people to achieve real ends in a real world, the analyst comes immediately into collision with a set of difficulties. Firstly, there is the fact that there is no one-to-one correspondence between linguistic form and function - that is, there is no one way to criticise; and hence the relationship between a speech act and the language used to realise it, is not neatly definable and incontrovertible. Further, just as no one form may be linked exclusively with any one function, so also more than one function may be achieved, at the same time, by any one particular form (Fraser, 1975; Richards & Schmidt, 1983a): a supervisor's inquiry may, for example, also be a criticism.

Other difficulties exist. A speech act is not necessarily co-extensive with an utterance unit or a part thereof; and therefore decisions about speech act assignment cannot be made in advance, independent of a knowledge of 'co-text' - defined as 'the stretch of discourse in which a particular utterance is embedded' (Leech & Thomas, 1988:42). As well, speech act labels tend to be under-descriptive in terms of the complexity of functions that utterances can achieve, as Franck (1981) shows in the example of the label 'statement of agreement' for a listener's minimal responses within an interaction.

Further, characterising illocutionary force from the exclusive perspective of speaker intention takes no account of how language is taken up and interpreted by the hearer, who, significantly in the next move, becomes the speaker (Goffman, 1981; Streeck, 1980): in the case of criticism, there is the instance when the speaker voices a criticism which is not interpreted as such by the hearer; or conversely, when an utterance is interpreted by the hearer as criticism when it was not intended as such. Streeck (1980:145) writes:

> What counts in the realm of human communication is not the acts as they are intended by speakers, but the consequences; that is, how they are interpreted and subsequently responded to by other participants.

Indeed, how one is heard 'is totally in the hands (or ears) of the hearer (Fraser & Nolen, 1981:96). In a similar vein, Labov and Fanshel reject the notion of conversation as 'a chain of utterances' in favour of the notion of a 'matrix' bound together by participants' responses and interpretations (1977:30). Edmondson (1981b) resolves the difficulty by distinguishing between a speech act and an interactional act, where the latter accounts for how a speech act operates within a discourse context.

Thus, despite the fact that forms of words do not carry their pragmatic value around with them (Crichton, J. 1993, pers. comm., 20 June), speech act theory does not offer the analyst a way of linking utterance with context or assigning meaning to form (Levinson, 1980, cited by Brown & Yule, 1983a:233). There is little account of how context is inextricably bound up with the interpretation of meaning; nor how meaning is to be retrieved from context and harnessed to the interpretation of the typically elliptical utterances of natural conversation (Franck, 1981). As well, the nature of context as 'social activity' means that there is no such thing as text-invariant rules of interpretation (Levinson, 1979). As well, there is the failure to recognise the range of functions a speech act may realise in respect to different strata of discourse - a view which would require the recognition that lower units are functionally connected to higher units within which they are embedded and to which they are connected in the meaning-making process (Streeck, 1980).

A further difficulty in speech act theory is its 'fait accompli' approach: its failure to deal adequately with the dynamic and strategic features of natural interaction. In this regard, Frank writes (1981: 229):

> By cutting a piece of conversation into speech acts... one does not reveal... the internal logic of the development of the conversation vis the strategies of the participants to steer and anticipate this development. It is based on a fixed *post factum* language perspective not on the continuously 'moving' perspective of the participants towards the unfolding structures of communication.

Meaning, then, is a negotiable commodity, being co-constructed over time by participants, being embedded in the unfolding nature of discourse, and being ever in the process of becoming - 'infinitely deferred' (Gilbert, 1992:45). This is linked, too, with the facility of

deliberate imprecision that language offers its users: any particular speech act is not yet able to be labelled thus at the point of utterance; indeed it may well shift to something else mid-stream; or be taken up by the hearer as something different from what the speaker intended. This is not to say that speakers cannot speak with precision; but rather that should they choose to be imprecise (Stubbs, 1983a; Channell, 1994), then the language amply and flexibly accommodates them. This facility for slippery indeterminacy pervades natural language, and needs to be seen as a 'constitutive', not a 'defective' feature of interaction (Streeck, 1980:148).

4.2.1.2.2 Text and context

The above comments testify equally to the complexity of natural language and to the inadequacy of speech act models of dialogue; and underscore the difficulty of the researcher's task (Orletti, 1984; Potter & Wetherell, 1987). Brown and Hoffman, for example, wrote of their process of interpreting supervisory discourse: 'in every instance coding is from the coder's viewpoint with meanings inferred from the verbal behaviour of the conference participants (1969:103). One is left with the unavoidable dimension of interpretation which, as Stubbs writes, is largely dependent on 'idiosyncratic or social knowledge', tempered by the awareness that 'we can never learn everything about anything' (Stubbs, 1983a:175).

However, meaning is dependent on 'neither God nor Humpty Dumpty' (Yule, 1985:91); and such knowledge that is brought to bear in the researcher's interpretation of language, is itself not insubstantial, embracing a close familiarity with both text and context. In terms of text, there is familiarity with four enveloping 'nests' of text: the immediate site of the supposed act of criticising; the 'sequential environment' (Streeck, 1980:142), including what precedes and what ensues the act; the slightly larger unit of the critical incident (explored further, below); and the even larger unit of the whole encounter or SD.

The researcher also brings to the task a familiarity with context. There is a considerable literature on the what is involved in such contextual knowledge (Cook, 1990). Levinson (1979), for example, writes of the activity type conventions that predict and constrain text. He defines these 'social episodes' (1979:393) as 'goal-defined, socially constituted, bounded events with constraints on participants, setting and so on, but above all on the kinds of allowable contributions' (1979:368). These constraints on what will count as allowable also help to determine how what one says will be taken i.e. they presuppose 'activity-specific rules of inference' (1979:393).

Others, too, have contributed insights. Goffman writes of the 'ritual' (1981:124) of certain kinds of speech events and the framing (1986) which facilitates comprehension and interpretation. Halliday states: 'a text is the product of its environment, and it functions in that environment' (1978:136). It follows that a knowledge of the roles and obligations and institutional settings in which a type of discourse typically occurs is needed in the assignment of meaning. Patterns that occur in any one supervisory discourse are not likely to be unique or idiosyncratic; but are rather likely to be comparable to those of other texts produced in comparable settings for comparable purposes. Goffman comments on the cosmos-building, game-like quality (1961:26-7):

> A matrix of possible events and a cast of roles through whose enactment the events occur constitute together a field of fateful dramatic action, a plane of being, an engine of meaning, a world in itself, different from all other worlds except the ones generated when the same game is played at other time.

The very fact that the SD is an identifiable genre - a 'recurring occasion' (Zimmerman, 1984:224) - well recognised in the literature (Hoover, O'Shea & Carroll, 1988; Lindsey, 1969; Rust, 1988; Waite, 1990b) indicates that there are shared interactional mores and purpose-driven objectives to which interactants orient themselves. As Gervasio wrote of the Labov and Fanshel's (1977) speech event of psychotherapy, supervision is a 'routinised form of verbal behaviour, delineated by well defined boundaries and sets of expected behaviours' (1987:109). Grenfell, for example, points to the script of a recurring socio-drama (1993:15). Pointing to evidence of such universality, O'Neal and Edwards' (1983) large-scale study of supervisory conferences concludes that 'although each participant... brings unique characteristics to the context of his/her clinical setting, certain themes appear to pervade all experiences (1983:40). Thus the interpretation of cues embedded in text relies on a close familiarity with the generic potential being played out through oft-repeated instances of a given activity type.

Familiarity with such features of text and context means that in the interpretation of utterances as mitigating, the researcher calls on resources comparable to 'members' resources' used in the construction of meaning (Fairclough, 1989:24; c.f. Goffman's 'realised resources' 1961:28; and a speech community's members' knowledge of rhetorical conventions, Linde, 1988:385). Indeed, such resources are indispensable interpretive tools, for no utterance type is intrinsically mitigating, but rather derives its politeness implications from its function, or meaning in context (Tracy & Eisenberg, 1990/91). Meanings assigned to speech acts are therefore derived from their implementation in social practice. This is a view adopted by Brown and Yule

(1983a), who steer a compromise path through the extreme positions of, on one side, considering language 'in isolation from communicative contexts' and on the other, examining each discourse fragment as individual or idiosyncratic.

4.2.1.2.3 Interpretation

In order to capture the dynamic nature of unfolding discourse, the approach taken to the recognition of utterances is essentially interpretive. Franck writes: 'since a speaker always has to anticipate the interpretation to be expected and the investigator necessarily has to adopt the position of a hearer, the heuristics of the investigator must start from the interpretation side' (1981:234). To make sense of the constructed meaning of 'two-person talk' (Goffman, 1981; Sanford & Garrod, 1981), the investigator must penetrate the encircling membrane and sense-make from within. Because researchers have no ready means of inspecting the consciousness of the speaker, they must impute speaker intent from the flow of talk, most especially from the hearer's follow-up utterance, which, as Goffman's analysis of 'replies and responses' shows, can be quite illuminating (Goffman, 1981; Gronn, 1981). As Lanigan suggests (1977, cited by Gronn, 1981:107), this is what listeners do as interactants:

> The listener must *supply* his [sic] own content: that is, the intent he would have if he were making the utterance in the role of speaker in the same situation (emphasis added).

Moreover, the supply of content is achieved through the interpretation of cues:

> Inter-subjectivity is empirically grounded in the reading of cues, cognitions and consequences: it does not involve reading other men's [sic] minds telepathically, but rather their communications (Ball, 1972, cited by Gronn, 1981:107).

Thus the researcher, like the member participants, is intimately involved in the construction of meaning, if not, of course, as it dynamically unfolds, then later, on reflection of the transcript. Like participants, the researcher must activate the speaker-hearer interpretative principles that facilitate the assignment of meaning to social events (Cicourel, 1969). Indeed, as Hudson details, hearers need three kinds of knowledge: knowledge of the constraints on the use of language; knowledge of the constraints on conversation; and knowledge of the universe, especially the speaker, and the preceding discourse (1975:4). The researcher, then,

is cast rather in the role of Goffman's 'bystander' (1981:132) who overhears and makes sense of what happens, but who is not a 'ratified participant' (1981:131).

While clues embedded in public meaning are rendered defeasible through an informed reading of text in context, what one arrives at can never achieve irrefutability, remaining 'a best guess' (Franck, 1981:234), a version of events, an account 'of what is being oriented to by the members' (Hester, 1981, cited by Gronn, 1981:107). Interpretation remains probabilistic.

4.2.1.3 The emergence of the critical incident

4.2.1.3.1 Defining the critical incident

Even while the decision was made to limit the functional analysis to critical FTAs, it was evident that this minimal unit was too small for the purpose of catching the cluster of mitigating redressive strategies that seem to adhere to FTAs. As Ch. 5 will illustrate, the effect of mitigation can be realised by the language of the FTA or by preceding language or by subsequent language, or by any combination thereof (Holtgraves, 1992). Thus something larger was needed, though not as large and unwieldy as the SD. Given that a FTA rarely happens entirely in isolation and that meaning is constructed dynamically and collaboratively (Brown & Yule, 1983a) as 'joint production' (Stubbs, 1983a:21), the attention now turned to the contiguous discourse - language adjoining and surrounding the FTA - that is, to its 'co-text' (Brown & Yule, 1983a:46), defined as 'the stretch of discourse in which a particular utterance is embedded (Leech & Thomas, 1988:42).

A preliminary analysis of the data displayed a repetitive pattern: a particular, albeit loose, configuration of moves clustering around FTAs. This configuration has three parts: a preamble or lead-up to the FTA; the FTA itself; and a denouement, consisting of a number of responses of the teacher and supervisor. The chunk so identified was called 'a critical incident' (CI) (App. 7). Thus:

	CI	=	P + FTA + CT + D
where	CI	=	critical incident
	P	=	preamble
	FTA	=	face-threatening act
	CT	=	co-text
	D	=	denouement

The CI of the present study conforms broadly to Nuthall and Lawrence's definition of an incident used in the analysis of teaching: 'any question or demand and all the subsequent verbal moves which occur up to, and including the final response to that question or demand' (1965:21). If one substitutes, in this quotation, 'critical FTA' for 'question or demand', the definition holds for this study. Corresponding to this study's preamble and denouement, Nuthall and Lawrence include introductory comments preceding the act and terminal comments following the last response. This accords with Held's structural schema: preparatory phase + focal phase + final phase (1989:191); and with Stech's (1979) notion of a 'sequence', being a series of turns pertaining to the same topic. It also neatly conforms with the structure of bad news messages (Holt, 1993): preparing/getting ready; delivery; and wrapping-up (Clark & LaBeff, 1982) or 'shoring' (McClenahen & Lofland, 1976)[1].

The CI was not always neat and obliging: naturally occurring language often resists the efforts of the analyst to fit it with off-the-shelf clothes. There is a murky grey area between the analyst's need for a unit of analysis and the power of natural language to defy such a unit as artificial and inappropriate. (It is perhaps part of the fascination of language that we keep trying). Like others, Brown and Gilman lament the difficulties or working with 'naturalistic' data: 'you do not control the flow of data. It pours over you and you must cope as best you can... data sets are often critically incomplete; analyses cannot be fully objective' (1989:208).

Variations on the idealised model took many forms. Sometimes the co-text was co-extensive with the FTA, or even with the CI. Sometimes there was no preamble, with the supervisor launching straight into the FTA (e.g. [1.2]). Sometimes, one FTA rolled into another almost without warning, threatening to hurl a spanner into the analyst's works, creating the analytical dilemma: to separate each point into its own CI or to keep all of them within the one (e.g. [7.1]; [7.2]; [7.3]). Sometimes the FTA was lurking, almost incognito, buried in a subordinate clause and almost missed by all parties - speaker, hearer and analyst (e.g. [6.1]). Sometimes an FTA took a number of forms, typically beginning quite indirectly, and building up towards a more explicit utterance (e.g. [9.2])[2], and the steps of its evolution provided the substantial part of the CI, rather than a follow-up denouement. At other times, the unit stopped short at

[1] The literature on bad news messages suffers from being restricted largely to professional-stranger dyads and based on recall or participant observation, not recorded incidents (Holt, 1993; Maynard, 1989a, cited by Holt, 1993).

[2] This conforms with Schegloff's observation that the bringer of bad news may not be the actual teller (1988a)

the end of the FTA with no subsequent discussion (e.g. [3.6]). Computing meaning from silence is onerous, as a range of optional interpretations present themselves (Pomerantz, 1984a): the teacher may have taken the point on board without discussion; or may have given a non-verbal acknowledgment cue; or may have chosen not to respond verbally; or may have been quietly appraising the situation; or may have misunderstood what was said; or a combination of some of these may have occurred[1].

Ideally, the CI was co-extensive with a topic of criticism, but this was not always the case. In one CI [10:4], three minor, substantially unrelated, critical remarks are made under a broader rubric. Similarly, in another CI [4:9], three related sub-topics are treated within the same critical incident. However, conversely, in another SD, three CIs cover the same critical topic ([4:1]; [4:2]; [4:3]). In the last case, the division into three CIs with each covering a sub-element of the topic is one born of the analyst's convenience: the desire to make unwieldy data somewhat more manageable.

Initially, such deviations, discrepancies and variations on the idealised CI were worrisome, with the researcher concerned to find an invariant unit that would fit, glove-like, on every empirical instance. After some experimentation, the realisation dawned that, given 'the complexity of de-limiting speech act boundaries (Pearson, 1988:76), any unit would inevitably generate its own discrepancies; and that this is evidence of the complexity of language rather than the sloppiness of the unit or the imprecision of the analysis. The unit of the CI is therefore somewhat idealised into its three parts, but recognition of its idealised nature does not prevent one from using it as a convenient means of getting closer to the data. Furthermore, the researcher took heart from the fact that discourse is inherently episodic (Gumperz, 1972) and that segments are generally recognisable, often through the retrospective cues afforded by the achievement of goals (Conte, 1981; Penman, 1990).

There are two 'down' sides to such a tight focus (Fleet, 1993), each of which in its own way compromises the 'gestalt'. Firstly, with the CI as the unit, the focus is on a component of the dialogue, not the SD in its entirety. Secondly, the approach taken means that the investigation is analysing supervisors' (not supervisees') language. With regard to the latter point, there is no doubt that a study of supervisee language is a fertile area, as Waite (i.p.) has shown and

[1] One means of accessing such invisible decision-making processes is through stimulated recall in a *post factum* interview, an instance of which is included in Ch.5.

as Pomerantz's (1984a) investigation of pursuits of responses suggests. That, however, is another project. These comments notwithstanding, the CIs that form the heart of the present research are nonetheless 'natural' in the sense that they occur without the intervention of the investigator (Fleet, 1993). Settling on them might be considered inevitable concomitants to heeding Bogdan and Biklen's advice (1982) on the need for narrowing a study to manageable proportions.

While speech act analysis necessarily involves the discourse context - how speech acts combine to form a system - the focus of the present inquiry is on utterance-level mitigation, rather than the structural or sequential properties of larger speech episodes of criticism. Nonetheless, it certainly would be a worthwhile project to determine how the actual sequential delivery of criticism operates: what elements are mandatory, what optional; what constraints exist; and what rituals and formulae pertain (Hatch, 1992) in the tradition of work that has been done, for example, with compliments (Wolfson, 1981), proposals (Houtkoop-Streenstra, 1990), political apologies (Abadi, 1990), and trouble-telling talk (Jefferson, 1984). This, too, is another project.

The present inquiry adopts the view underpinning the discourse work of Sinclair and Coulthard (1975:13):

> The level of language function in which we are centrally interested is... the level of the function of a particular utterance, in a particular social situation and at a particular place in a sequence, as a specific contribution to a developing discourse.

The concern is with the componential structure of the FTA identified as a recurring pattern in supervisor talk, rather than with stretches of discourse. The analogy may be made with still photography vis-a-vis moving film: while the entire conference is perceived as a dynamic unfolding event, the focus of this inquiry is on the 'stills' lifted from the moving film. These snapshots are frozen, single moments in time - slices of discourse at the text-context nexus captured at the moment of the FTA.

By way of summary, the audit trail to this point has shown how the CI, despite the 'furry' bits, became the criterial unit by which the data bank was assembled. The pilot data [SD1] was carved up into its 10 CIs, and these, numbered chronologically, were re-compiled into a new data base. Then with subsequent SDs, the same procedure was followed: first, the SD was

combed for its critical FTAs; then environmental parameters - the boundaries of the CI - were determined; and then the CIs were numbered chronologically within their SD.

4.2.1.3.2 The meaning of 'critical' in the critical incident

The use of the term 'critical incident' involves a fortuitous conflating of five meanings of 'critical'. These do not include the use made in some educational research literature (Andersson & Nilsson, 1964; Drachelet, 1981; Flanagan, 1954; Gordon, 1973; Housego & Boldt, 1985), where the usage is much broader[1]. These meanings are 'unpacked' in the discussion below.

- Bearing criticism

The first meaning operates at the surface level of the FTA and refers to the supervisor's act of criticising the teacher. Typically, the supervisor says or implies that something could/should/might have been done differently/better. This coincides with Kimsey's (1969) behavioural categories for supervisor talk, namely evaluation and prescription. Given that supervision involves bridging the gap between what is and what ought to be (Cogan, 1973, cited by Housego & Boldt, 1985), criticising may be thought of as central to the supervisor's role and obligations.

- Key events

The second meaning of critical is the sense of important or key events (akin to the usage in the supervision literature). The moments in the discourse where the supervisor criticises, remarks on a weakness, targets an aspect of teaching worthy of improvement are certainly important components of the SD, although of course supervision involves more than this (Turney et al. 1990).

[1] The difference in fact is two-fold. Within the context of educational research, the CI takes in the whole experience of the practicum, whereas the case boundaries of this study are the speech event of the SD. Secondly, the educational usage of 'critical' means events which are perceived by key protagonists to be important either positively or negatively, within the parameters of a certain experience, such as the practicum (e.g. 'war stories' in Burgess & Briscoe, 1993:10); whereas in this study 'critical', in its most immediate, referential sense, means 'bearing criticism'.

- Delicate condition

The third connotation is the sense of tenuous or touch-and-go, as in the 'critical' condition of a hospital patient. Here the allusion is to the fact that the SD as a communication event is neither facile nor robust; and may be constituted of some parts that are more prone to conflict or breakdown than others. These elements, it is argued here, are the CIs, where face threat is at its height, and where the supervisor might be thought of as 'treading on eggshells'. It is at such points that the tension in the supervisor (between protecting the face of the other and delivering a clear critical message) is most palpable. In these moments, the supervisor has to navigate a delicate path: raising the teacher's awareness of difficulties without deflating morale. Juggling these tensions requires not inconsiderable dexterity; and because the process is far from fool-proof, it is fraught with risk and susceptible to breakdown.

There is a deeper implication, as well, in which typical 'encounter difficulties' (CIs in the present study) are viewed as the expected locus of action, where one might justifiably expect particular strategies to be mobilised in order to protect or salvage the communication. Candlin (1987) sees particular merit in 'explaining moments of conflict in discourse', for it is at such sites that one might expect maximum energy to be given over to repair, recovery and salvage. Choices made in discourse - from the 'meaning potential' of language (Halliday, 1978:39) - may be very telling about a matrix of interlocking elements - the user, interlocutor, topic, setting, situation and purposes. In J. Thompson's rather cryptic words 'discourse says something about something' (1984:137): The CIs of the present study may be viewed as junctures of this type, for as already mentioned, critical FTAs seem to generate within their own co-text a cluster of strategic, redressive actions (Brown & Levinson, 1978) - the very substance of this investigation.

- An analytical process

The fourth sense in which critical is used is a perceptual one and relates to the sampling procedures used in the research methodology. Here critical means analytical - looking closely at something normally taken for granted, or seeking 'to unfamiliarise the familiar' (Burns, 1992:109). By scrutinising the FTA, the researcher is able to focus on the features of communication that are mostly highly relevant for this study. The intention linked with the point made above about sites of difficulty, is to avoid the routine in search of the problematical. This is an 'estrangement technique' that allows us to render visible (through

'making strange') what we do not normally perceive (Stubbs, 1983a:243). It allows us to 'see with different eyes' by becoming 'exiles from the familiar' (Plessner, 1978, cited by Burns, 1992:109).

- De-naturalisation

The fifth sense of critical is related to critical linguistics (CL), an approach which views language as 'a (or indeed the) major locus of ideology' (Fairclough, 1989:12). This entails investigating verbal interactions 'with an eye to their determination by and their effects on, social structures' (Fairclough, 1985:747). It allows one to speak of institutional meanings rather than personal meanings, viewing those in institutional roles (such as supervisors) within institutional rather than personal frames of reference (Crichton, J. 1993, pers. comm. 2 May). 'Critical' here is intimately connected with ideology: a governing purpose of CL is the process of 'denaturalising', or rendering transparent the opacity that accompanies taken-for-grantedness.

4.2.1.3.3 Analysing the critical incident

Once the SD was carved up into its constituent CIs, the CIs were themselves analysed. This process fell into two phases, the first structural, a matter of displaying the internal organisation of each CI; the second, functional, having to do with identifying mitigation.

The display of the internal organisation of each CI involved numbering the CI into its internal sequential moves. In this study, the term 'turn' (Sacks et al. 1974) is used interchangeably with 'move' (Sinclair, Forsyth, Coulthard & Asby, 1972), and indicates the stretch of speech that ends when another speaker begins. Each turn or move is equivalent to that chunk of language which Brown and Hoffman called an utterance unit 'the uninterrupted verbal behaviour of either participant in a supervisory conference' (1969:102). Derived from Bales' (1951) thought unit, this also served as the basis for later studies (Zeichner & Liston, 1985; Zeichner et al. 1988). However, in this study, the term 'utterance' is used differently, as discussed below.

The process of demarcating a speaker's turn is not without difficulties, as evidenced, for example, by the dilemma of how to interpret backchannel cues typical of listener behaviour (Sands, 1988), or what to make of attempts at turns that are thwarted, perhaps by being

unsuccessful interruptions or by being themselves interrupted. For ease of reference, a notation of three numerals within square brackets is used to refer to the data: thus [3.4.6] refers to the third SD; and within this, to the fourth critical incident; and to the sixth turn within this. Altogether 10 SDs were collected and analysed, yielding in total 96 CIs. One SD had as many as 15 Cis; and one as few as one; and the mean was 9.6.

Once the structural organisation was displayed, the functional analysis followed: the supervisor's turns were scrutinised for the phenomenon of mitigation at the level of utterance. By 'utterance' is meant that basic unit of spoken language that may or may not be co-extensive with a turn or move, but that generally is co-extensive with a speech act or function (Hymes, 1972b; Labov, 1972b). The utterance, then, is defined in functional terms, and is not a grammatical concept, to be equated with a clause or sentence or a combination of these. In other words, the utterance is closely bound up with the nature of spoken language. James refers to it as a discourse or interactional unit measurable with reference to semantic, pragmatic and phonological correlates of 'completeness' (1983:196). Widdowson (1978; 1979b), for example, distinguishes between written and spoken language with related pairs of terms: usage/use; sentence/utterance; locution/illocution; text/discourse; cohesion/coherence. This is compatible with Levinson's pragmatic distinction (which follows Bar-Hillel, 1971):

> A sentence is an abstract theoretical entity defined within a theory of grammar, while an utterance is the *issuance* of a sentence, a sentence-analogue or sentence-fragment, in an actual context (1983:18, emphasis added).

The question that guided the search for evidence of mitigation was: what is happening in the supervisor's language (specifically in and around the FTA) that achieves the effect of softening or hedging the criticism? As the organisation of the data into various categories is discovered through or generated from the data itself, the approach may be termed 'emic' (Kreckel, 1981; Pike, 1954; Wolcott, 1988). While an etic analysis is based upon existing 'outside' universals, an emic approach locates the phenomena under investigation with reference to 'its own internal system' (Waite, 1990b:42). This allows the researcher to ask and answer the question 'what is happening?, rather than the 'etic' question 'Is X happening?' as is the case when the coding categories are imported into and imposed on a study. Given the nature of this research as illuminative - metaphorically, lighting a candle in a dark room - an emic approach was considered highly appropriate (Parlett & Hamilton, 1977).

As utterances and elements of utterances were recognised as mitigating, they were analysed linguistically. Chapter 5 is devoted to a full description of these, but some preliminary samples may serve here[1]. Quite obvious as mitigators were minimisers ('just', 'a little bit'), the use of vague language ('sort of' 'kind of'), and modal verbs ('could', 'might') which all, in different ways, operate to take the edge off the harshness of what is being said. Less obvious, perhaps, are syntactic devices, such as the avoidance of agency - instead of saying 'you failed to drill the target language', the supervisor might refer to the fact through nominalisation ('the drilling'). As well, there is deictic usage that exploits tense (e.g. through distancing time to the past) and aspect (e.g. using the durative aspect to make an action seem less sharp) for mitigating purposes. These and other elements by which mitigation is achieved are fully discussed in Ch. 5, which forms the 'heart' of the investigation.

4.2.2 Grounded theory

4.2.2.1 Generating a data-driven typology of mitigation

There was a need to classify the mitigation into a schema that would impose a sense-making order upon the evidence. Recognising an instance of the discourse as mitigating was a preliminary step to generating such a typology. The development of an empirically-grounded typology of mitigation became the re-formulated goal. As can be seen thus far in the audit trail, and as is typical of qualitative research, progress within the investigation involved 'progressive focusing' (Parlett & Hamilton, 1977:15; Stubbs, 1983a:231).

To achieve this, a number of steps had to be carried out, some of them simultaneous: data collection; identification of critical FTAs and the boundaries of the CIs; transcription of the tapes; structural and functional analysis of the Cis leading to the recognition, interpretation and classification of mitigators . These are discussed under 'research issues', below. It will be clear, from the foregoing discussion, that operating here, within the data collection and processing mechanisms, there are three different units. The largest is the SD, the unit of collection, the bounded speech event recorded in its entirety. The smallest is the FTA, a functional unit, the speech act of criticism delivered by supervisor to teacher. The middle one is the CI, a structural unit, its core constituted of FTA + co-text.

[1] App. 8 provides a micro-analysis of how mitigation is interpreted pragmatically.

To generate an exhaustive and entirely data-driven typology of mitigation, a grounded theory approach was adopted (Glaser & Strauss, 1967), involving the following steps. First, the SD, through the constituent CIs, was analysed for evidence of mitigation. As each instance of mitigation was identified, it was classified as a type (or sub-type) of mitigator. Each classification forced the investigator to re-consider the categories already in existence, in order to decide whether the new instance could be incorporated into an already designated class; or whether a new category had to be created to accommodate it. The classification of each instance, then, meant a re-appraisal of previously classified instances.

Categories of the emerging typology were derived, therefore, by a process of constant comparison: comparing the latest instance with previous ones to see whether the emergent typology can currently accommodate it or whether it needs to be modified. The constant re-appraisal generates an ongoing process of amendment and refinement which itself allows the emergent typology to becomes steadily more robust. Accordingly, when all the CIs of the first SD were accounted for within the typology, the next SD was processed in the same way, and the next and the next. The process - by which the researcher inductively created her own categories, emic-style - requires both convergent and divergent thinking (Guba & Lincoln, 1983), being not unlike Lincoln and Guba's (1985) index card system with its categorisation organised according to look/feel-alike qualities.

4.2.2.2 Procedural orderliness

Hindsight gives the procedure of constant comparison a certain linear neatness which the actual processes lacks at the time of analysis. Neat in the reporting, the procedure is inherently elusive being, by definition, unstable and constantly in flux. Other procedures were put in place to allow the concurrence of collection and analysis to be kept reasonably orderly; and to keep the investigator's thinking both critically focussed yet creatively buoyant - twin, competing and seemingly contradictory impulses in the qualitative researcher. Guidance from the research literature was sought (Bogdan & Biklen, 1982; Merriam, 1991); and decisions made accordingly e.g. narrowing the study (i.e. selecting the criticism-specific FTA; honing down the SD to the CI as the unit of analysis); reformulating analytical questions as required (e.g. the movement from 'fragility' to 'restraint' to 'mitigation'); using memos and observational comments throughout; carrying out 'member checks' (Guba & Lincoln, 1983:316); ongoing reading of the literature; creative playing with metaphors, especially in the naming of

categories, as an aide-memoire as well as a cure for the near-sightedness that plagues the close-up, intensive data processing that is required here.

In fleshing out and stabilising the categories, guidance was also gained from the application of some of Holsti's (1969) criteria, borrowed from content analysis. In the first place, categories had to continue to reflect the purpose of the research, guarding against the tendency to become side-tracked. This meant bearing constantly in mind that the study was in search of an explanation for the fragility of the SD: thus the search for mitigation was for evidence of restraint. Secondly, categories had to be derived from a single classification principle: in this study, this was furnished by the initial definition of mitigation, which served well as a stable point of reference. However, as the discussion of the typology in Ch. 5 will demonstrate, the nature of language defies the creation of mutually exclusive categories, for overlap is both considerable and inevitable.

4.2.2.3 Saturation

The grounded theory approach cannot predict how much data will be required. The criterion for judging when to stop sampling is what Glaser and Strauss call 'theoretical saturation' (1967:61). At the level of the category, this means

> that no additional data are being found whereby the (researcher) can develop properties of the category. As he [sic] sees similar instances over and over again, the researcher becomes empirically confident that a category is saturated (1967:61).

While in theory data collection can continue indefinitely - there is always another dialogue that could be taped, transcribed, and analysed - the researcher has judiciously to select an end-point. Lincoln and Guba's realistically grounded definition is the one chosen by this investigator: saturation is reached when 'continuing data collection produces tiny increments of new information in comparison to the effort expended to get them' (1985:350). Stabilisation is a gradual slowing down. By the end of the tenth supervisory dialogue, following the identification and analysis of 96 CIs, and accounting for the creation of 14 classes within the typology, the investigator was satisfied that the typology had stabilised to accommodate the data in which it was grounded and that further efforts in data collection were likely to be poorly rewarded. Ch. 5 involves a description and pragmatic account of each class and sub-class of the typology of mitigation.

4.2.2.4 Applying the typology: intensive analysis

The next stage was intensive analysis. This involved a return to the 10 SDs, and a re-analysis of the evidence of mitigation in the light of the stabilised typology. In this second 'sweeping' of the data, each instance of mitigation was re-located and coded according to its classification in the typology.

During the first stage of the analysis, as the typology had been in the process of emerging, it had not been possible to fix securely upon any classification decision, as the process of constant comparison meant that everything was constantly subject to re-interpretation. However, once the typology had emerged as stabilised, it was then possible to *apply* the typology as a criterion-referenced instrument.

Thus, the first phase of the analysis was essentially emic - using categories evolved inductively from the data - while, the second seems more etic-like, in that, in the consideration of each instance, one is asking: which class or sub-class in the typology does this instance fall into? However, even this could not be considered truly imported, for in a truly etic approach 'the categories for coding the data are developed in isolation from the behavioural system and are imposed on the data, while the characteristics of the data are still unknown to the researcher' (Gibb, 1990:3). Thus, even while phase two of the analysis seems one step removed from the data through the mediation of the typology, the approach is nonetheless essentially emic, as the typology is itself firmly grounded in the data.

4.2.3 Research issues

This section covers some of the research ground mentioned, but not discussed in depth, above. It fills out the details of the research audit trail.

4.2.3.1 Data collection

4.2.3.1.1 Early intentions and constraints

The original intention of the researcher was to attach herself to a short-course adult TESOL training program, such as the one she herself established at the Institute of Languages,

University of New South Wales (UNSWIL) in the early 1980s, for it is with this type of course that she gained most, though not all, of her experience of pre-service teacher education[1]. Most significantly it was in this context that the experience of feedback-as-fragile was grounded.

The intention had been to make the study course-specific, as in a qualitative case study approach (Merriam, 1991), using ethnographic tools to gain a holistic picture of supervisory practice, covering a range of participant perspectives. Knowing that the language was to be scrutinised closely, the intention was to heed Flanders' (1976) warning about the dangers of de-contextualised data analysis; and therefore to embed the research in its natural context. The aim was for the researcher to immerse herself fully in the research context, through the modality of non-participant observation, so as to better implement a pragmatic interpretation of text.

Rarely, though, does reality match the ideal; and therein lies the spawning ground for compromise. The researcher was unable to gain the needed institutional access, and so an alternative approach was sought - using the SD[2], rather than the whole course, as the case boundary[3].

4.2.3.1.2 The case boundary

Thus it was that the SD - as much a process as a speech event - became the 'instance' around which the inquiry was to focus (Adelman, Jenkins, & Kemmis, 1983:2). One attraction was the apparent 'boundedness' of the SD in time and space as a speech event. This, however, is not to suggest that the conference is de-contextualised, or viewed as an abstract unit unconnected to the experiential or psychological realities of concurrent, prevailing and historical issues. Sufficient is known about the social ecology of situations (Goffman, 1972a; Ball, 1973) for no

---

[1] These courses were conducted by the researcher at UNSWIL (1982-87) and as a private consultant for other teacher training centres (1988-1991).

[2] The defining characteristics of the SD are detailed in 4.2.3.1.3, below.

[3] Obstacles sometimes fortuitously become opportunities. With hindsight, it seems that, even in an atmosphere of trust, non-participant observation is too intrusive an instrument to be of great value. While it offers insight into extra-lingual features of the interaction (Waite, 1990b; Waite, i.p.), the proximity of a third person-observer in a dyadic genre is discomforting.

under-estimation to be made here; and certainly, the ecological case for interpreting the conference as 'unbounded' has been made eloquently by Waite (1990b; 1992b; i.p.). Included within the ecological context-embeddedness of the conference, are participants' cognitive domains: a complex array of knowledge, experience, assumptions, expectations (Cook, 1990) that shape and permeate the boundaries of the conference; and that provide members with 'scripts' (Brown & Yule, 1983a:240) and 'resources' (Fairclough, 1989:11) which serve them in the making and interpreting of meaning.

Yet, as Goffman shows, the crucial defining ingredient of focussed encounters is 'the participants' maintenance of continuous engrossment in the official focus of activity' (1966:11). This engrossment sets up a 'selective inattention' (Sullivan, 1956, cited by Goffman, 1961:38), determining what is, and is not, dwelled upon. This engaging activity acts as an 'encircling barricade' (1966:25-6) - 'a boundary around the participants, sealing them off from many potential worlds of meaning and action' (1966:25). In Goffman's (1961) view, the activity that defines the encounter is world-building: it generates a set of meanings that accrue to it (1961:27). Yet the barricade is screen-like, selecting and transforming what it lets in; and determining its own rules of relevance (1966).

Thus the SD might be considered both bounded and unbounded. It is bounded in the sense that the dynamics of the encounter are 'tied to the functioning of the boundary-maintaining mechanisms that cut the encounter off selectively from wider worlds' (Goffman, 1961:66). It is unbounded in the sense that there are elements in the external milieu which will shape and constrain the meanings of the encounter. Certainly, from the perspective of research method, the location of the SD in time and space provides a convenient reference point, affording the type of characteristics conducive to case study. As Guba and Lincoln (1983:86) put it:

> The boundary problem comes down to this: How is the inquirer to set limits to his [sic] inquiry? What are the rules of inclusion and exclusion? How can the inquirer know what is relevant and what is not relevant?

As is outlined below, the SD affords ready answers to these important questions.

4.2.3.1.3 The supervisory dialogue defined

Data collection then proceeded on the basis of the following four defining characteristics of the SD:

- It had to be a face-to-face encounter between two people[1].

- The two people had to be in the institutional roles of supervisor and student (or neophyte) teacher i.e. there had to be some asymmetry between them. In TESOL practice this mostly means dyads of the following types: trainer or supervisor with student teacher; co-operating teacher with student teacher; teacher development co-ordinator with neophyte. It excludes the increasingly common modality of peer feedback, where two teachers co-operatively engage in each other's own development through mutual observation feedback cycles (Wajnryb, 1993a)[2].

- The meeting had to take place after an observed lesson or lessons. This required the supervisor to be familiar with the teacher's teaching, so helping to ensure (though not guarantee) that the agenda was teaching-specific (see next point, below). It excluded other supervisory meetings, such as pre-lesson planning conferences (Acheson & Gall, 1987); or 'general progress' meetings.

- The agenda of the conference had to deal with details of actual observed teaching. This guarantees that the supervisor will have observed, if not the whole lesson, then a sufficiently large section of it to be able to comment specifically on the teaching. Data were not collected from more general, non-teaching-specific meetings which were less likely to reap a rich harvest of critical FTAs, as the supervisor is more able in such a situation to walk the un-conflicted path of helper/support-giver. In contrast, the teaching-specific SD seemed most suitable for yielding critical FTAs because it is the point where supervisors give feedback about teaching. This is not to suggest that they cannot do it elsewhere or that they must do it here, but rather that it was most likely to happen here.

4.2.3.1.4 Sampling

Sampling decisions involve questions of site (where?), time (when?), subjects or participants (who?) and events (what?) (Burgess, 1982:76). In harmony with the nature of qualitative case study, where answers are sought not to questions like 'how much?' or 'how often' but rather to provide insights into human experiences, processes and relationships (Parlett & Hamilton, 1977), the most appropriate strategy is non-probability, purposeful sampling (Patton, 1980).

---

[1] Precluded from use, therefore, was the modality of group feedback, characteristic of the RSA/UCLES Certificate in TEFLA.

[2] One such peer conference had been used in the initial pilot study, as a means of examining the discourse patterns generated by relational symmetry. This issue, however, was not pursued further.

The particular purposeful sampling strategy used is criterion-based sampling,involving a 'recipe of attributes' (Goetz & Lecompte, 1984:77): the above four-point definition of the SD was used as a criterion for the selection of data. The design included sequential, theoretical sampling, involving an ongoing sample selection process. The process, by nature evolutionary, involves simultaneous efforts in data collection and analysis; and is inevitably guided by the emergent theory (Burgess, 1982; Glaser & Strauss, 1967; Merriam, 1991).

4.2.3.1.5 Data description

The data, 10 dialogues in all, derive from TESOL teacher education sites; and are displayed in Table 2. Contextual information pertinent to specific dialogues is contained in App. 9.

SD#	1	2	3	4	5	6	7	8	9	10	11
SD1	N	F	+	F	M	A	T/TE	Pr	Syd	1988	V
SD2	S	V	-	F	F	H	T	Pr	Per	1992	A
SD3	S	F	-	F	F	A	T/TE	Pu	Syd	1984	V
SD4	E	M	+	M	F	A	T	Pr	Syd	1992	A
SD5	N	V	+	F	M	A	T	Pr	Syd	1992	A
SD6	S	M	+	F	F	H	T	Pr	Per	1992	A
SD7	S	V	+	M	F	A	T/TE	Pr	SF	1992	A
SD8	N	F	+	M	M	A	T/TE	Pr	Syd	1988	V
SD9	N	F	+	F	M	A	T/TE	Pr	Syd	1988	V
SD10	S	V	-	M	M	H	T	Pu	Syd	1991	A

Table 2: Descriptive details of the 10 SDs in the grounded theory study

Key:
1: Status of teacher: S = student; N = neophyte; E = experienced
2: Supervisor experience: F = few (<10 conferences); M = many (10-30); V = very experienced (>30).
3: Supervisor training: + = supervisor has training; - = supervisor has no training
4 & 5: Sex: F = female; M = male.
6, 7 & 8 : Institutional context:
A = adult learning context
H = senior high school
T = language teaching context
T/TE = Teaching and teacher education context

Pr = Private
Pu = Public
9: Site: Syd = Sydney, Per = Perth, SF = San Francisco
10: Time: date (year) of conference
11: Recording mechanism: V= video; A = Audio

Most of the sites are concerned with adult teacher training, a few with upper secondary school contexts (which in most elements of ESL methodology is not radically distinct from adult-oriented programs). Some of the sites are centres devoted to language teaching; while others also include a teacher training program. Most of the Sds involve student teachers, some neophytes (less than a year's classroom experience), and in one case, a teacher of about three years' experience. In this last case, though, the teacher was new to the school and new to teaching in Australia, so there was about her an element of the 'neophyte', as reflected in the tenor of the conference.

The degree of professional supervisory experience of the supervisors varied from 'nil' to 'very'. In regard to training, very little supervisory preparation exists in Australia, particularly in this field (hence one motivation for this study); and where training is noted (yes/no), it is in the form of a short training course (30 hours) of which about 25% was devoted to conferencing/feedback skills.

All except one of the 20 participants - 10 supervisors and 10 teachers - are native-speakers of English. The one exception, a francophone, had near-native speaker proficiency. As a general rule, the researcher deliberately excluded non-native speakers because of the potential influence of pragmatic cross-cultural failure (Thomas, 1983).

Though gender is not an issue under investigation in this study, it is recognised that mitigation has been linked with women's speech (Lakoff, 1976; Brown, 1980). To avoid a gender-influenced distortion, the sampling covers the full range of possible gender interactions: male supervisor/female teacher (x2); female supervisor/male teacher (x3); male supervisor/male teacher (x2); female supervisor/female teacher (x3)[1].

The data are also drawn from a range of institutional contexts, both public and private language teacher training schools and centres, largely from the Sydney area, but two come

[1] Because the TESOL trainee population as well as TESOL supervisor population in Australia is female-intensive, a particular effort had to made to gain a spread of gender through the sampling.

from Perth, and one from San Francisco. The data were collected over an eight-year span, using either video and audio tapes, some of which were originally recorded for purposes of supervisor training[1].

The above description of the specifics of the data corpus illustrates how aspects of Denzin's

Within *time* triangulation, there are data from different dyads at different times through an eight year period. While this does not fit snugly into Denzin's cross-sectional or longitudinal sub-categories, it does nonetheless suggest greater validity than would be offered by data totally collected from one point in time (such as, for example, was envisaged in the original plan for a single course-specific case study).

Space triangulation attempts to counter the effects of parochialism in studies exclusive to country or sub-culture. The multi-sited nature of the study - calling on a range of settings across institutional, urban and national boundaries - increases the robustness of the study.

Subject triangulation is sought through calling on supervisors whose diversity - in respect to training, experience, gender - affords a validity to the data that is greater than that which would be offered in a more subject-homogenous study.

Exhibit 3: Time, space and subject triangulation

(1970) triangulation typology have been used to provide internal validity to the investigation (outside of the between-method triangulation discussed at the start of this chapter). These aspects include time, space and subject triangulation; and are detailed in Exhibit 3.

4.2.3.2 Transcription

As the data were collected, the CIs were identified, and then transcribed. This was done in large part by the researcher but also involved a research assistant. Where the assistant was employed, he keyed in the first draft of the transcription; this was then submitted to a second, third and fourth 'run' by the researcher. Employing a research assistant not only shares the intensely labour-consumptive side of producing verbatim records; it also counters threats to validity in having the data exposed to more than one set of ears. Validity was further increased by the multiple drafting process and by returning transcription samples along with the

[1] As ethically required, all participants gave permission, either at the time of the recording or subsequently, for the material to be used for stipulated research purposes (see App. 11).

recordings to some of the supervisors (50%) for their comments. These yielded insubstantial changes, giving the researcher confidence that another investigator transcribing the dialogue for the same purpose would arrive at a roughly similar product. Returning transcripts to speakers, rather than, say, exposing them to another 'outside' rater as a validity check, is an endorsement of a participant's own interpretation of events (Stubbs, 1981; 1983a), a tradition more honoured in sociology than linguistics, which has rather tended to denigrate user's accounts and insights as naive and unreliable (Martyna, 1980).

Transcribing involves the transformation of a temporally organised speech event to a spatially represented text (Gronn, 1981). This shift in channel from an aural-meant-to-be-heard to a written-meant-to-be-read medium is radical, entailing complexities that have been amply discussed (Bull, 1989; Kress, 1979; Ochs, 1975; Sacks et al. 1974; Stubbs, 1983a). The process is also laborious, time-consuming and costly.

Yet for an investigation such as the present one, transcription is indispensable. Its importance has been outlined by Van Lier (1988): it by-passes the hazards of real-time coding; it provides the necessary means of estrangement; it allows the researcher the facility of immersion; it provides for multiple exposures so that the apparently trivial may surrender its significance; it allows users and other researchers to analyse the same data, and, with regard to the latter, affords a comparison with other investigations, thereby enabling cumulative research. As Moerman points out, 'we build our experienced, lived in, significant social reality out of a mesh of interactive processes too tiny and too quick for the thinking, planning "I" to handle' (1988:30).

These advantages notwithstanding, the process is fraught and tenuous, not least because a range of cues (verbal, non-verbal and paralinguistic) occur to some degree simultaneously (Ball, 1972; Gronn, 1981), and such issues warrant some discussion[1]. Firstly, even with the researcher's best attempt to represent in writing what the ear hears (supported by the eye in some cases, where the dialogue was videoed), the concession must be made that pure objectivity is elusive. As multiple interpretations may in theory be generated from the one acoustic signal (Brown & Yule, 1983b:xi), 'there is... no single correct transcription for a given utterance' (Stubbs, 1983a:229). Related to this is the fact that agreement may be reached over

[1] The thorny issue of context and its transcription (Cook, 1990) is acknowledged but not tackled, if only for the attempt to keep the task finite.

what is heard but not over what is meant[1]. Between the signal and its representation stands the mediating human transcriber, who inevitably and necessarily interprets what is heard.

Secondly, because all transcription systems are 'necessarily selective' (Heritage & Atkinson, 1984:12) - 'transcription is a selective process reflecting theoretical goals and definitions' (Ochs, 1979:44) - the inescapable truth is that product reflects purpose (Cook, 1990; Fleet & Cambourne, 1989; Stubbs, 1983). The analyst's purpose in transcribing the data mediates in the transcribing process; and such selectivity is inevitably reflected in the product - the actual transcription that emerges. In other words, transcription is itself an interpretation (Kress, 1979). The present study falls within the tradition of research in educational supervision (Lindsey, 1969) by focussing on the actual verbal language used by participants - 'the vocalised language used to communicate meaning' (Lindsey, 1969:118). As a consequence, signals that are non-verbal have been excluded as peripheral. Excepted therefore is information related to phonology - intonation, tempo, rhythm, volume, pitch, voice quality; as well as paralinguistic information pertaining to facial gesture, gaze, body orientation and other elements of proxemics and kinesics; as well as reference to the inter-relationship between talk and gesture (Jefferson, 1984). Likewise, too, intricate fine details in the delivery of spoken language - e.g. fine pausing, alignment, latching, silences and sub-types of interruption - are overlooked, apart from the rather gross categories of short and long pauses (App. 10).

This is not to say that such features do not signal meaning (Hall, 1959; Gronn, 1981, and references contained therein; Gumperz, 1977; Gumperz, 1990; Willing, 1992). Indeed, there have been systems, albeit complicated and fallible, that have been worked out for their transcription (Bull, 1989, and references contained therein; Goodwin, 1981, and references contained therein; Kendon, 1990). The abiding principle here, however, is that the researcher's orientation reflects the research purpose: what for this researcher is peripheral, defined by the research purpose, may well be central for other researchers, defined by their purposes[2].

---

[1] There is, for example, the case of 'uh huh' and its equivalents, which have been variously interpreted as 'passing moves' (Clark & Haviland, 1977); 'minimal responses' (Fishman, 1978; Zimmerman & West, 1975); 'passive strategies for being civilly egocentric' (Derber, 1979); and 'backchannel' or 'listener responses' (Duncan, 1972; Rosenfeld & Hancks, 1980).

[2] So, for example, in a study on male-female discourse (Tannen, 1990), one may expect close attention to be given to latching and interruptions, warranting a classification akin to the sensitive binary flow chart developed by Roger, Bull & Smith, (1988); and in a study of non-verbal communication, one might expect gaze, gesture and body orientation to be of central concern (Atkinson & Heritage, 1984; Haddad, 1991; Kendon, 1990; Prutting & Kirchner, 1983, and references contained therein). Further, on the basis of cross-cultural data differences, Hirokawa and Luebs (1994) argue against the notion of

As this study is bound up with meaning embedded in verbal language and its context of situation, it is to some extent possible to 'tidy up' the apparent chaotic appearance of natural conversation, to regularise the orthography, and present transcript records that are reasonably pleasing to the eye and not 'perceptually unreal' (Stubbs, 1983a:229). Notwithstanding the danger of standard orthography resulting in 'masking' (Ochs, 1979:45), the advantage of readability should not be understated. Gronn warns that the 'straining for accuracy in rendering the naturalness of speech' may 'clutter the text, taxing both the reader's concentration and patience' (1981:109). Likewise, Burns refers to an 'unvarnished transcript' as 'impenetrably mystifying' (1992:322).

While supra-segmental phonological phenomena have not been transcribed because they are largely not related to the research purpose, it must be said that other reasons exist too for the avoidance of this less transcribable element. The first, already stated, is that it is in words that the main research interest lies. Here Labov and Fanshel's criticism of Scheflen's (1973) quantitative study for having moved away from words as the key element, has been heeded: 'the primary data for the listener are the words being spoken by the... parties, and whatever interpretation he (sic] constructs will be based upon them (1977:21).

Secondly, the researcher was not sufficiently convinced that there exists an agreed-upon notation system that can satisfactorily and validly represent non-verbal acoustic signals. Fear of being drawn into a kind of 'folk phonetics' (Stubbs, 1983a:229) fuelled the reluctance to explore this path of investigation. Furthermore, the elusiveness of a precise prosodic coding system is connected to the very nature of language which seems to defy or resist the process. That this is inevitable and not simply a lamentable feature of the channel of speaking (Labov & Fanshel, 1977), will be explored further in Ch. 5. Suffice it here to say that there is a need for language to be 'slippery' (Carr, 1993:10), even evasive, for speakers on occasion need communication to be deniable: vagueness affords us a crucial 'retreatability' without which we would doubtless threaten, more than we already do, a great deal of interlocutor face. Stubbs contends that 'without such imprecision, life in the social world would be impossible' (1983a:174). This elusive quality means that the problem of the notational interpretation of both prosodic and paralinguistic cues may well be beyond resolution.

---

a standardised transcription system.

A third and related point is that because the paralinguistic dimension is largely sub-conscious, we have 'an extreme paucity of metalinguistic terms' to describe such features (Willing, 1992:5), which itself reinforces the problems of awareness and analysis[1]. While the present study largely neglects the non-verbal dimension, the natural redundancy of language means that prosodic cues are usually supplemented by more definable (and transcribable) features (Willing, 1992). Thus, while prosodic and kinesic means of indicating mitigation no doubt exist in isolation (Brown & Levinson, 1978), they also very often exist redundantly, with the mitigation being cued both verbally and non-verbally. This investigation focuses solely on linguistic means of signalling mitigation.

Nonetheless, one may regret the non-representation of prosodic features on two counts. Firstly, there is no question that making sense of spoken language require a piecing together of non-verbal acoustic signals. Stubbs uses the spelling 'co*hear*ance' as a reminder 'that many of the linking mechanisms of conversation require to be heard to be appreciated' (1983:19). Secondly, there is no doubt whatever that mitigation can be achieved through a modifying intonational influence (Labov & Fanshel, 1977), among other prosodic cues. Hatch has shown, for example, that when compliments are used to soften criticisms, they are often accompanied by a special intonational curve that signals the forthcoming 'but' (1992:139). In the data of this study, too, the supervisor who said *'that dialogue that you used in fact was r-e-a-s-o-n-ably long... wasn't it?' [3.1.1]* signalled her mitigation partly through the elongation of the first two syllables of 'reasonably'. At other times *sotto voce*, a quickening of tempo or a rising tone similarly achieve a mitigating effect. However, for the reasons outlined above, this study remains a word-based, utterance-level investigation of mitigation, and hence concentrates on the transcribed spoken word. As with other researchers (Gronn, 1981; Grimmett & Crehan, 1990; Roberts, 1990), the transcription has been standardised to conform broadly with conventions of written English.

4.2.3.3 The observer's paradox

During data collection, another concern bearing on the validity of the study relates to what Labov (1972b) termed the observer's paradox: the notion that the very act of observing (or

[1] It is perhaps not surprising that it is the field of developmental pragmatics, with its close attention to the verbal and non-verbal aspects of child interactions, that insights are being offered into the transcribing of such features (see Ochs, 1979 and references therein).

listening or recording) may affect the interaction, yielding contaminated data, and giving rise to the question - what language is 'natural' language?

While the observer's paradox is widely known, and has a certain intuitive 'rightness', it is notoriously difficult to identify and equally as difficult to guard against. Furthermore, as Wilson (1987) points out in his discussion of data as a product of methodology, there is not a great deal of evidence supporting the relative success of various avoidance methods. The difficulties are formidable. For one thing, evidence of reduced observer's affects can only be produced comparatively, by contrasting affected with non-affected data (Wilson, 1987). This assumes that we already know what is 'unaffected'; but if we are unable to observe without contaminating the evidence, how can we pretend to know what is unaffected? In this light, it is not difficult to see what led Wolfson (1976) to argue that natural speech *per se* does not exist; what exists is 'speech appropriate to a particular contextual moment of production' (cited by Wilson, 1987:161). Weimann (1981) explored the issue of reactivity to recording procedures by examining four levels of obtrusiveness on a continuum from overt to covert. McLaughlin (1984) speculates, along with Weimann (1981), that observer effects are likely to be greater in naturalistic than in controlled laboratory settings where a pre-existing arousal may subsume any subsequent reactivity to recording equipment. However, she concludes that there is 'no convincing evidence on this point one way or another' (1984:244).

Nonetheless, if claims are being generated from the data analysis, it is important to protect their validity through 'sensitivity to the problematic of the paradox' (Wilson, 1987:161), since only such sensitivity will allow one to militate against observer effects. In regard to such counters to threats to validity, five points warrant mention here.

The first point relates to the formality-informality continuum of early Labovian research into speech styles. Labov (1972b) argued that the more formal the speech situation, the greater the conscious attention speakers gave to their speech. Such a premise led to the assertion that the informal, vernacular style - displaying the least conscious attention - would be the style against which other styles would be calibrated. In the present study, certain factors - the institutional setting, the power asymmetry of roles, the generic staging and 'pre-programmed' organisation (e.g. turns, topics) that characterise formal registers - dictate that the speech style or register of the SD be considered of formal orientation. This is relevant to the issue of observer effects: as participants, perforce the genre, are already paying careful attention to their language (in particular, the supervisor), it would be a formidable task indeed to distinguish between those

elements that derived from attention-to-speech, generated by the speech style, and those generated though observer effects.

A second point is methodological, having to do with the role of the researcher vis-a-vis the speech community being studied. In regard to this question of 'audience', Labov's 'sociolinguistic interview' gained fame as a means of accessing the vernacular for study purposes. For example, following a Labovian approach, Milroy (1980, reported in Stubbs, 1983a and Wilson, 1987) developed a similar stance: researcher-as-audience in the status of 'friend of a friend'. In the case of the present research, the supervisor dialogues were solicited from supervisors with whom the researcher had developed long-standing relations of mutual respect and trust - in some cases, over ten years. Some of the supervisors were trained as teachers, then as trainers, by the researcher herself and thus had covered a great deal of professional 'territory' together. If Labov is right about the inevitability of observer's paradox, then Bell's hypothesis (1984, reported in Wilson, 1987) of a principled theory of audience design may be a strategic solution. In the present study, the researcher-as-audience is acknowledged as an unavoidable influence and the contention is that the shared membership of the speech community and a shared history of professional contact reduce the seriousness of observer effect.

The third point relates to the assertion made by Labov (1978), supported by Milroy (1980), and to a lesser extent by Wilson (1987), that observer effects are more salient at the start of a recording than at later points. In Weimann's (1981) study, for example, a decline in measured anxiety over the course of an interaction, suggested that the obtrusive effects may be dissipated over time. Indeed, the researcher's previous experience of taping and transcribing learners (Wajnryb, 1988b) supports the speculation of decreasing reactivity over time. Fortuitously for the present study, the 96 CIs occur in the latter part of their SDs. This, of course, is not coincidental: supervisors tend, indeed are advised (Turney et al. 1990) to lay a groundwork of positive support before initiating points of criticism. This means that CIs tend not to be drawn from the early stage of the SD and as such, one may speculate, are less prone to observer effects.

The fourth point relates to the nature of potential distortion in contaminated data. Wilson's (1987) typology of contaminated speech includes four categories: direct tape-affected speech, in which explicit reference is made to the observation or recording; personalised tape-affected speech, in which the recording equipment becomes 'a ratified participant' (Wolfson,

1976:200); formal tape-affected speech where instances of distortion are caused by observer affects; and indirect tape-affected speech, where language is pragmatically motivated by an awareness of the 'researcher as audience' (Wilson, 1987:175)[1].

Applying Wilson's typology to the present data, there are two instances worthy of mention. In one instance [SD7], the supervisor makes oblique reference to the tape-recorder (Wilson's category 1); and in the second, the supervisor [SD2], early in the dialogue, treats the researcher as participant (Wilson's category 2), In neither of these cases does the occurrence occur *within* a CI (which is the only language being analysed for mitigation). There are no obviously discernible instances of Wilson's category 3 (but see the earlier point about the difficulty of discerning between formal-attention-to-speech affects and observer affects). Wilson's category 1 is the most difficult to identify, but in this regard, it is likely that distortions will be in the direction of cautious inhibition rather than openness. Indeed, two supervisors, responding to questions designed to probe this very issue, commented on their awareness of the recording equipment. One [SD4] said that it made him more tactful and indirect ('tentative', 'diplomatic') than he considers he usually is[2]. The other said the recording equipment made her more 'self-conscious' and 'guarded' than usual [SD1]. In the light of the fact that this study is seeking out evidence of mitigation, one might argue that distortion encouraging mitigation is certainly preferable to distortion suppressing it.

The last point relates to information provided to participants prior to recording. In the active effort to reduce observer's paradox, participants were not told that the study aimed to identify and classify mitigation, for this would have been to invite arousal and self-consciousness, and therefore court contamination. Rather, some minimal deception was involved in describing the study in non-specific and non-technical terms: 'an investigation into the roles of supervisor and teacher in the feedback context' (App. 11).

---

[1] The place of the invisible observer makes sense, too, within Goffman's (1981) framework of participant status.

[2] This was ratified by a subsequent viewer of the video who had experience of receiving feedback from that supervisor. The viewer claimed that the recording process had made the supervisor less 'blunt': 'he's usually more intimidating'.

4.2.3.4 Validity

Validity is a measure of the match-up between reality and the researcher's interpretation of reality. Establishing validity requires determining 'the extent to which conclusions effectively represent empirical reality' (LeCompte & Goetz, 1982:32). In the following discussion four potential sites of threat to validity will be discussed: the recording process; the transcribing process; the identification of FTAs; and the linguistic analysis of mitigation.

The first potential threat existed in the recording phase, and relates to the observer's paradox. One factor which reduces the paradox in this case is that recorders are not at all an alien feature of the language teacher's classroom, where most teachers are familiar with and comfortable users of both audio and video machines. This is not to say that such familiarity assures content validity (Helmstadter, 1966), but rather that it reduces the threat (Linde, 1988; Rundquist, 1992:433, and references contained therein). In addition half the recordings took place in the context of teacher education courses, where recordings are a frequently employed training procedure - e.g. in micro-teaching sessions and for purposes of demonstration lessons - and being on display and accountable may be said to be part of the territory of training (Wajnryb, 1990b). There was also the factor of trust in the personal relationship that shaped the bond between researcher and most of the supervisors and some of the teachers involved. While this may not overcome observer's paradox to the degree that Heath was able to do in her research (reported in Nunan, 1992a), trust is a crucial, albeit nebulous, factor in countering observation effects. Furthermore, post-recording interviews of subjects, designed to elicit their impression of the effect of the taping, generated two comments of import (discussed earlier).

The second site of potential threat was the transcription: here the aim was to reduce the gap between the mechanical voice recording of the 10 SDs and the written representation of these dialogues. As discussed earlier, the validity of the transcribing was shored up by having a minimum of four (and often more) drafts, and 'going through' two sets of ears: the researcher's and an employed research assistant. The transcripts are as-verbatim-as-possible replications in writing of what was said. The attention was placed on the replication of words, while prosodic and kinesic features were largely ignored. The decision to represent the dialogues in language that closely resembles normal orthography meant that other analysts - e.g. in the inter-rater reliability check - would be able to access the transcripts readily. Samples of the transcriptions, along with the original tapes, were returned to speakers who were asked to comment on the fidelity of the scripts. These emerged from the check with only minor adjustments.

A further potential site of threat is in the recognition of critical FTAs, an issue discussed in depth earlier. To check for threat here, the researcher's selection of the CIs was 'member checked' by having a sample returned to the supervisor who was asked to comment on the extent to which the CIs identified on the tapescript were indeed sites of criticism. As a substantial number of these contained pragmatically ambivalent utterances and were not at all straight-forwardly explicit critical acts, the check was warranted. The resultant correspondence between the researcher and the supervisor was 100%, giving the former confidence that her ability to recognise the CIs and demarcate SDs correctly was well founded.

The last danger area is in the linguistic analysis, touching both the nature of the data and the analysis thereof. Important considerations here are that the data, collected over a time span of eight years, are authentic, derive from naturalistic contexts, and express the exact language of participants. Indeed, it is one of the strengths of qualitative research that 'the lengthy data collection period provides opportunities for continual data analysis, comparison and corroboration to refine constructs and to ensure the match between research-based categories and participant reality' (McMillan & Schumacher, 1989:191). The time span also gave the researcher the opportunity to refine insights: most particularly it allowed an increasing familiarity with the constructs and tools of a pragmatic interpretation of language. As Gadamer states, 'we stand in traditions' (1972, cited by Van Manen, 1977:218), and these afford us our hermeneutic, interpretive frame. This gain in knowledge over the data collection/analysis period afforded the researcher increasing confidence in the emergent typology. Likewise it is a strength of qualitative research that it admits 'into the research frame the subjective experiences of both participants and investigator'. This has to be balanced against stringent self-monitoring, what Erickson terms 'disciplined subjectivity' (1973, cited in LeCompte & Goetz, 1982:47).

4.2.3.5 Reliability

A study may be said to have internal reliability 'if independent researchers, on analysing the primary data, come to the same conclusions as the original investigators' (Nunan, 1992a:60). The researcher in this study was keen to establish whether the emergent typology was able to be used by independent raters or observers in a manner comparable to the researcher. What was sought was both agreement on the categories and predictable assignment of data to

different categories (Nunan, 1992a). A key element in establishing internal reliability was the ability to replicate findings independently.

The first major difficulty here was the fact that, as with many of the concerns of qualitative research, we are not afforded the convenience of low-inference descriptors (LeCompte & Goetz, 1982). The area under scrutiny is inherently unobservable: the analyst has 'to infer unobservable mental states from observable behaviour' (Nunan, 1992a:60). In the case of the present study, on analysing the primary data, the rater has to assess speaker's intention from evidence of mitigation in the discourse; and this, as has been discussed earlier, requires a reading of clues-in-text, as well as a familiarity with the supervision activity type. In addition, the fact of the 'inherent ambiguity of the phenomenon being coded' (Penman, 1990:22) militates against the process of classification. Furthermore, given that much language use is 'naturalised' and difficult to see (Fairclough, 1989; Stubbs, 1983a), some experience with 'estrangement' procedures (e.g. discourse analysis, CA, text linguistics) is necessary before an analyst can see beyond surface visibility. The 'anthropological gaze' (Carr, 1993:10) is not easily acquired.

In order to apply the coding system on which the typology is based, a rater would have to be educated to see through and around the language and to make meaningful connections between text and context. Given that it had taken the researcher some fifteen months to reach a confident point in her own readings, training others to this point was indeed a challenge. There were also practical constraints, such as the time involved for independent raters to immerse themselves in the data. However, establishing adequate pre-coding training for the inter-rater reliability check in this study is comparable to the extensive time spent by multiple researchers on prior training and discussion during fieldwork to reach agreement on meaning (Becker, Greer & Hughes, 1968; Becker, Greer, Hughes & Strauss, 1961); and such efforts invested by qualitative researchers in reaching consensus on agreed meanings is a guard against threats to internal reliability and hallmark of good research. The human instrument is the qualitative counterpart of mechanical instruments and statistical procedures and 'can become more reliable through training and practice' (Merriam, 1991:171).

Substantial effort was therefore made in the recruitment of two suitable independent analysts and in their training. Regrettably, circumstances dictated that the two raters were trained

separately; and early in her training process, one rater had to withdraw. Practical constraints then led to the decision to pursue the training with one independent rater only[1].

The training process involved four steps. Firstly, the researcher gave a detailed explication of the typology through many illustrative samples drawn from the data. This was preceded by the preparation by the researcher of a provisional manual, which the rater had studied in advance of the first training meeting, ensuring that he arrived at the meeting already familiar with basic concepts. One aim here was to familiarise the rater with the coding system and with the logic with which samples were classified. A second aim, related to ensuring internal validity, was to seek agreement on the general constructs of the typology, and to test the researcher's intuitions and hunches against another's. According to LeCompte and Goetz, it is more appropriate to the paradigms of qualitative research to seek agreement 'on the description or composition of events rather than on the frequency of events' (1982:41). Accordingly, intensive discussion was carried out between the researcher and the rater, which gave assurance to the researcher that the intuitions underpinning her typology were well grounded.

The second step of the training process involved fabricated data for these served better as training tools. The third stage involved trial runs on authentic samples, with accompanying analysis and discussion. Here the aim was to introduce the rater to naturally-occurring linguistic complexity and the decision-making and priorities that have to be considered in coding. This stage is a necessary intermediary before moving on to the final phase, the test on authentic data. The final stage involved the rater assigning data instances to categories of the typology using authentic texts. On the test texts, instances of mitigation were marked with an empty box; the rater's task was to assign a code number to the box according to the typology (App. 12).

The training covered some seven hours of one-to-one contact, completed over two half-day sessions. The results, at the end of the seven hours, showed a correspondence of 94% between the researcher and the independent rater, well over the standard criterion of 75% (Fairhurst et al. 1984). It is an index of the high-inference nature of the descriptors that such intensive training was required to reach this correspondence. Despite the difficulty of the

[1] The rater, David Cervi, was chosen for both his academic and professional background: M.A. (Applied Linguistics), Macquarie University; substantial experience in teacher education, especially teacher supervision, which topic served as his own research investigation for his Masters thesis (Cervi, 1991); as well as for his willingness to participate.

descriptors, the result of the inter-rater reliability check suggested that the researcher might conclude that the typology was a relatively reliable instrument.

The issues of internal and external reliability are crucial research matters, and given the unavoidably high-inference nature of the descriptors in the present study, a great deal of attention has been given in this report (see Ch. 5) to accounting fully for the descriptors in the typology. In effect, an 'audit trail' of linguistic elements has been provided, allowing itself to be deployed as a de facto 'operating manual' (LeCompte & Goetz, 1982:40) for other researchers. This affords replicability, both internally, using independent raters, and externally, applied to other data, in similar and related fields[1].

Two of LeCompte and Goetz's suggestions for guarding against threats to internal reliability - low-inference descriptors and multiple researchers - were, regrettably, not able to be utilised in this study, the first because of the sphere of concern; the second because this is a single-researcher project. However, other advised strategies for protecting internal reliability have been utilised (LeCompte & Goetz, 1982; Lincoln & Guba, 1985; McMillan & Schumacher, 1989; Merriam, 1991; Nunan 1992a). One of these is peer examination which was operationalised in two ways: firstly, through cross-site validation, by incorporating insights gleaned by other researchers in other sites and studies (principally, Grimmett & Crehan, 1990; Pajak & Seyfarth, 1983; Roberts, 1990, 1991a; Waite, 1990b); and secondly, through peer review of published articles in established forums, incorporating samples of primary data (Wajnryb, 1993b; Wajnryb, 1993d; Wajnryb, i.p.). In addition, internal reliability is increased through the fact that the primary data are preserved in mechanically collected recordings (audio and video). Moreover, participant-researcher services were obtained through key informants in the ethnographic side of the study, allowing triangulation through corroboration of findings.

An important component of reliability is a clear statement of the investigator's position, embracing the guiding constructs and underpinning premises of the theoretical framework. This research offers such a statement in the first section of Ch. 5, where the conceptual

[1] For example, it may be of interest to other researchers to find out whether the patterns of mitigation reported here of TESOL teacher education are similar to those in other subject areas or other teaching contexts; or whether, as is the researcher's hunch, that they are indeed neither subject-specific, nor context-specific (primary, secondary, adult), but rather are a product of the interface between (any) supervisor and (any) teacher.

framework of pragmatics is extensively described. In addition, a careful audit trail of the research process - including criteria by which subjects were chosen and a description of these - is included in this chapter, its data displays and associated appendixes.

Overview

This marks the end of the discussion of research method in the linguistic study. The reader now has two options. The first option is to remain with the linguistic study, turning now to Ch. 5, and the findings of the linguistic research, after which a return may be made to the next section of Ch. 4 (research method in the ethnographic study). That is:

Ch. 4, first prong -> Ch. 5 -> Ch. 4, second prong -> Ch. 6.

The advantage of the above is the sustained focus it offers.

Alternatively, the reader may wish to complete the entire chapter (Ch. 4) on research method, before reading Ch. 5 on the linguistic findings and then Ch. 6 on the ethnographic findings i.e. following the usual chronology: Ch. 4 in full -> Ch. 5 -> Ch. 6.

4.3 The second prong: research method in the ethnographic study

> We do not experience language in isolation - if we did we would not recognise it as language - but always in relation to a scenario, some background of persons and actions and events from which the things which are said derive their meaning (Halliday, 1978:28).

The second prong of the research involved an ethnographic inquiry into supervisor attitudes, thinking and decision-making processes. The approach involved accessing 'participant meanings' (McMillan & Schumacher, 1989:405): discovering how supervisors conceive of their professional roles and duties; and how they themselves explain or 'make sense' of their experiences. In particular, the aim was to access their view of feedback - what they saw as the governing, shaping and constraining influences of that speech event - and their perspectives on delivering critical feedback. If the first prong set out to describe the verbal behaviour central to the inquiry, the second prong aimed to understand 'what that behaviour

means to the person under study' (Wiersma, 1991:220). Because qualitative methods are grounded in the realities of everyday life, they seem ideally suited to interpreting the semantics of human experience (Good & Good, 1981).

The second prong sought corroboration for the hypotheses that emerged from the empirical study of supervisor language in the primary study. The research method is outlined in this chapter, with the findings and discussion treated in depth in Ch. 6.

4.3.1 Data collection

4.3.1.1 Interviewing as a means of data elicitation

The research interview has been described as 'a conversation with a purpose' (Webb & Webb, 1932:130). More specifically (Cannell & Kahn, 1968, cited by Cohen & Manion, 1989:307-308), it has been defined as

> a two-person conversation initiated by the interviewer for the specific purpose of obtaining research-relevant information, and focused by him [sic] on content specified by research objectives of systematic description, prediction, or explanation.

The purpose in this study was to get inside the mind of supervisors, or, as Waite puts it (1990b) 'behind the other set of eyes', in order to see the feedback conference from their point of view: 'a reconstruction of the inner world' of the subjects' experience of feeding back in supervision (Hycner, 1985:291) - an 'exploration of the emic' (Adelman, 1981c:7). There have been many studies of supervisee perceptions of the conference (as outlined in the survey of perception studies in Ch. 2); fewer have been drawn from the perspective of the supervisor.

Because the information sought is invisible, mere observation of conferences, while yielding rich rewards in some regards (Waite, i.p.), would not answer the central questions: How do supervisors approach the conference? How do they interpret their role? What concerns shape their approach? What worries impinge and constrain them? What conflicts concern them? How important is face? What strategies do they deploy to facilitate their duties? What decision-making underpins their actions?

The purpose of interviewing 'is to allow us to enter in the other person's perspective' (Patton, 1980, cited by Merriam, 1991:72). Compared to using survey questionnaires (Cohen & Manion, 1989), interviewing is more time-consuming, inefficient and costly; suffers greater threats to reliability (bias and subjectivity); can cover fewer respondents; and is more dependent on subjective interpersonal factors of trust, rapport and questioning skill. Furthermore, semi-structured interviewing materials are difficult to pre-test and results are often non-aggregatable and unpredictable because of standardising difficulties (Guba & Lincoln, 1983).

The strength, however, is the capacity to elicit, probe and explore rich and deep insights into the world of thinking, feeling and decision-making of subjects. Semi-structured interviews are particularly valuable for eliciting the respondent's own frame of reference; there is less chance of misunderstandings; greater flexibility; more opportunities to probe, re-direct and deepen. They encourage 'portraits' of events in the respondents' natural language, enabling access to 'elites' - those with specialised knowledge of the situation under scrutiny (Guba & Lincoln, 1983). As such, the interview proved itself to be a worthy instrument.

The interview was chosen over other possible data elicitation mechanisms for its three primary purposes (Cohen & Manion, 1989) were closely linked to the needs of the study. Firstly, it allowed the gathering of information that had direct bearing on the research objectives. As Tuckman says (1972, cited by Cohen & Manion, 1989:309), it

> makes it possible to measure what a person knows (knowledge or information), ... likes or dislikes (values and preferences), and... thinks (attitudes and beliefs).

Secondly, the interview allowed the testing of hypotheses that emerged from the primary linguistic study. Thirdly, it is able to be used in conjunction with other methods in order to provide a rich, balanced and complementary picture of people and processes involved in the speech event under scrutiny. Thus, while the linguistic exploration afforded access to the language, and the experimental prong allowed the measurement of supervisee perceptions of such language, the ethnographic inquiry allows entry to the world seen through supervisory eyes.

The conception of the interview in this study (Kitwood, 1977, cited by Cohen & Manion, 1989:310-312) is essentially a naturalistic one - an encounter necessarily sharing many of the

features of every day life - rather than a material transaction involving information transmission and guarded-against bias. Within this conception (promoted by Cicourel, 1964), natural constraints of every day life shape, interact with and fill out the 'objective' message. Interviews in this sense are primarily and inescapably social encounters.

4.3.1.2 The focussed interview

Having decided on the worthiness of the interview as a data elicitation mechanism, the next step was to determine which of the various kinds of research interview was best suited to this project. These issues largely concern degree of structure and interviewer control (Cohen & Manion, 1989; McMillan & Schumacher, 1989; Merriam, 1991; Guba & Lincoln, 1983).

An interview may be characterised by two main factors: the extent to which one can pose *a priori* questions; and the extent to which one does or does not know in advance what it is one does not know (Guba & Lincoln, 1983) In Guba and Lincoln's (1983:159) graphic representation, there are two domains of knowledge - propositional and tacit - which may be mapped onto a range of interview type (characterised by their questions) in order to locate and determine the appropriate format. This study, as explored below, employs a schedule that falls firmly in the structured/propositional rather than the unstructured/tacit domain. Questions are able to be structured in advance, based as they are on stated hypotheses (see below). As well, what is not known is able to be formulated in advance: it will be largely related to alternative explanations to the face explanation of mitigation in supervisory discourse. Following the Venn diagram (Guba & Lincoln, 1983:159), the situation in this study is one of 'partial ignorance of structured situation'.

The choice was largely framed by the fact that the ethnographic study was part of the triangulation protocol of the research. In other words, its main purpose was to corroborate, through a distinct methodology, the hypotheses already established by the primary study. For this main reason, the interview format settled on was the focussed interview (see below), reported first by Merton and Kendall (1946) and described by Cohen and Manion (1989).

The focussed interview may be compared to other well known interview formats. In some ways (though not all), it resembles the 'elite' interview (Guba & Lincoln, 1983:166), in that the subjects have specialised knowledge; the topic requires in-depth dealing; there is an interest in etiology and in uncovering respondent motivations, intentions, explanations and

meanings. However, in one important sense - the corroborating rather than exploratory nature of the inquiry - the focussed interview differs from the elite version.

The focussed interview also differs from the ethnographic interview (Spradley, 1979) in two primary ways: firstly, with most respondents there was only one interview, whereas Spradley used a multiple interviews with a core of key informants; and secondly, the focussed interview is confirmatory in intent, as distinct from the exploratory intent of Spradley's ethnographic interviews. Nonetheless, some core features of Spradley's ethnography were deployed: the meeting between the participants is characterised by its explicit purpose; a range of ethnographic questions; and the inclusion of ethnographic explanations, especially about the project, the taping and the terms of reference (Spradley, 1979).

Owing to the 'insider status' of the researcher, there was some concern in regard to the language of respondents' explanations, Since the goal of ethnography is 'to describe a culture in its own terms' (Spradley, 1979:59), the researcher ideally would have aimed to have informants use 'native language' explanations rather than activate their 'translation competence' (1979:19). This, of course, works best when the informant speaks within his/her own culture and provides rich explanations for the researcher-as-outsider. In the case of the present inquiry, however, the researcher already had established insider status. This meant that many elements that might have been explained as if to an outsider fell into the category of 'assumed cultural knowledge'. It required considerable conscious effort on the part of the researcher to foster the production of unassumed ethnographic descriptions. On the 'plus' side, the shared professional background of researcher and informant meant that points of reference were easily established, and interviews could cut to the heart of the matter relatively quickly.

The choice of the focussed interview derived from the precise match between the situation at hand and the four criterion uses of this mechanism (Merton & Kendall, 1946; Cohen & Manion, 1989) These relate to issues of sampling; hypothesis verification; the construction of an interview guide; and the focus on subjective experience. These are discussed below.

4.3.1.2.1 Sampling

The first criterion is that the subjects are known to have been involved 'in a particular concrete situation' (Merton & Kendall, 1946:541) - here, the supervisory experience of giving feedback to a teacher following an observed lesson. The present inquiry used purposeful (or criterion-based) sampling to select supervisors. The particular sampling strategy adopted was typical case selection (McMillan & Schumacher, 1989; Merriam, 1991). A profile of characteristics possessed by an average case was created: a teacher who had had supervisory experience with pre- or in-service teachers, involving lesson observation and post-lesson feedback conferencing. Following this, instances of the case were sought.

However, the typical case sampling strategy was influenced by other sampling strategies. As is typical of ethnographic research, selection and sampling strategies are viewed as dynamic and concurrent, rather than static and *a priori* (McMillan & Schumacher 1989). What in fact began as typical-case selection, in practice blended with 'network' selection (McMillan & Schumacher, 1989:397), where each successive participant is named by a preceding one as fitting the profile of attributes. While this was not pre-determined, systematic snowballing, it emerged naturalistically (e.g. 'have you spoken to X? - she would be an interesting person for you to interview'); and thus inevitably influenced (and enriched) the final selection. Furthermore, because in this case the researcher is an 'insider', she was able to call on her own wide network of professional contacts; thus the sampling over-lapped somewhat with the category of 'reputational' case selection (McMillan & Schumacher, 1989:397).

Of the 30 supervisors sampled, some emerged as key informants: individuals 'who have special knowledge, status or communication skills who are willing to share that knowledge and skill with the researcher' (McMillan & Schumacher, 1989:406-407). Disproportionate time is rightly invested in such key informants (Wiersma, 1991:230) so as to tap their knowledge and wisdom. In such cases, the interviews tended to be longer, more penetrating and in some cases involved follow-up or multiple sessions. In the case of two informants who lived a long distance from the researcher, follow-ups were conducted by letter and fax, typically with questions eliciting answers, which themselves generated more (deeper, probing) questions, generating an on-going and highly insightful dialogue between researcher and respondent.

4.3.1.2.2 *A priori* hypotheses

The second criterion needed for focussed interviews is the existence of *a priori* hypotheses. In that such exist, they make this study unusual for ethnography which is typically thought of as hypothesis-generating, not hypothesis-verifying (LeCompte & Goetz, 1982; Nunan, 1992a). It must be remembered, however, that the ethnographic prong is part of the triangulation of a primary study, which itself generated a set of hypotheses for which corroboration is sought.

Thus, the researcher begins the ethnographic phase somewhat less in the dark than she began the linguistic phase. What was exploratory in the linguistic phase, now becomes more confirmatory: 'elements in the situation which the researcher deems significant have previously been analysed... (S)he has thus arrived at a set of hypotheses relating to the meaning and effects of specified elements' (Cohen & Manion, 1989:326). The hypotheses that emerged from

---

The researcher seeks, from supervisors, corroboration:

1 that the delivery of criticism is perceived and experienced by them as a face threatening act;

2 that criticism, when delivered unmitigated, poses a threat to morale and is counter-productive to supervisory goals;

3 that they perceive a potential danger to the clarity of the critical message, given the pressure to attend to face;

4 that they perceive and experience an intra-personal role conflict in their duties: between 'the helper' and 'the critic' which may translate into competing pressures - face-oriented versus message-oriented;

5 that mitigation is seen as face-motivated, a way of cushioning bad news, of hedging the force of the criticism but at the same time not opting out of delivering the message;

6 that mitigation may be problematic: too little indicating lack of support; too much creating obscurity in the message, suspicion etc;

7 that they perceive the need for strategic rather than social skills in the management of the face-to-face components of the encounter and recommend/welcome such strategic training.

Exhibit **4**: Corroboration

the linguistic study (the counterpart in this study of Merton & Kendall's preliminary content analysis) and that frame the interview are displayed in Exhibit 4 as a series of seven points on which corroboration is sought.

4.3.1.2.3 Interview schedule

The third criterion for the focussed interview is the hypothesis-driven construction of an interview schedule (App. 13). Foreknowledge reduces the investigator's task 'since the interview need not be devoted to discovering the objective [sic] nature of the situation' (Merton & Kendall, 1946:541). The difference between a variable-driven schedule (Cohen & Manion, 1989:321) and a hypothesis-driven one must necessarily be in the degree of open-endedness of the questions: in the former, the researcher seeks answers to open-ended questions; in the latter she seeks confirmation (or otherwise) of relatively close-ended questions. Merton and Kendall (1946) recommend that the questions that make up the focussed interview be guided by a set of four closely connected concepts: non-direction, specificity, range and depth (App. 14).

4.3.1.2.4 Subjective focus

The fourth criterion for the focussed interview is the spotlight on subjective experience. These are 'bounced off' the pre-existing hypotheses so as to test the validity of the latter. There must also be room for unanticipated responses, allowing for the generation of new, unforeseen hypotheses. This involves an open-endedness, a willingness to re-think, re-appraise and change direction on the basis of new incoming data.

4.3.1.3 Paradox: informant insights vs naturalisation

Central to the ethnography is the notion of participant meanings, itself the source of a fundamental paradox. On the one hand, the research gives credence and a prominent role to user knowledge (Stubbs, 1981); on the other, there is the awareness of the constraints of 'naturalisation' - i.e. the taken-for-granted obviousness that people ascribe to their automatic daily behaviours, language included. Fairclough describes how subject positions, situation types, meanings of words and interactional routines are all naturalised and perceived as normal, obvious, conventional and non-ideological - 'seen as simply "there" in a common sense way, rather than socially put there' (1992:9). Lakoff and Johnson (1980) have shown

how metaphor is embedded in our natural language, yet functions below the conscious level of awareness. In regard to supervision practices, Garman claims that we have, at a subliminal level, 'accepted the ritual as literal', and having thus accommodated ourselves, we now fail to see the insidious nature of daily activities (1986c:150).

Quite apart from the ideological issues at the core of critical linguistics, the concern here is with the fact that users may or may not have conscious or complete knowledge of the language they use (Braun, 1988; Martyna, 1980); nor of the motivations that underpin their usage. If one wishes to tap into participant meanings but these are unable to be articulated, there is a difficulty. Related to this is the danger that the process of 'raising respondent awareness' may interact with elicited responses and so generate a Hawthorne effect (McMillan & Schumacher, 1989:309).

The solution to these difficulties was to build an amended stimulated recall technique (Nunan, 1992a) into the interview that allowed respondents to generate their own authentic-like language[1] and then analyse it. This allowed them to recognise (and own!) mitigation almost at the point of spontaneous generation. While this retrospective elicitation procedure suffers from the same threats to validity as other elicitation and simulation devices (although perhaps less because the recall follows immediately upon the simulation), it seems to be a reasonable compromise in the light of the paradox mentioned earlier.

4.3.1.4 Data Records

Interviews were audio-taped using a small, unobtrusive cassette recorder with a conference microphone which allowed comparative comfort and movement. The advantages of audio-taping - completeness of the verbal interaction; freedom of the interviewer to concentrate on the interaction in real-time; opportunity for later re-checking; as well as reliability checks - were off-set against the chances of mechanical failure and respondent suspicion.

Notwithstanding the existence of a prior hypotheses, an ambience of 'conversation' not 'examination' was encouraged so as to facilitate fuller and freer responses (following Willms,

[1] Thus, at some appropriate point, the interviewer would say: 'imagine I am a teacher whom you have observed and you are now feeding back to me and there is an aspect of my teaching that you want to critique ... (at this point the supervisor-respondent nominates a topic)... now feedback to me on this point'.

Best, Taylor, Gilbert, Wilson, Lindsay & Singer, 1990). While questions were necessarily structured, responses were allowed to lead where respondents wanted, with minimum interference. Interviews lasted 30-40 minutes.

The 'interview log' (Merriam, 1991:84) was used to document the interviews (App. 15)[1]. Processing took place shortly after occurrence so that contextual features were still in memory; and elaborations (McMillan & Schumacher, 1989:411), where relevant, were included on the logs. The log captured the counter number and main points made by the respondent, including relevant, explicit and 'rich' quotations. During the preliminary analysis stage (see below), the logs were used for creating codes and gleaning key words.

As a check on validity, samples (20%) of the logs were returned to informants who were asked to comment on the fidelity of the document as a record of the interview (App. 16). The logs thus validated emerged satisfactorily from this fidelity check with only minor and insubstantial comments added.

The study then involved three broadly consecutive stages: preliminary data analysis (combining both collection and on-going analysis); intensive data analysis; and report-writing. The first two of these are discussed below; the report is Ch. 6.

4.3.2 Preliminary Data Analysis: coding

4.3.2.1 Procedural simultaneity

Qualitative research does not offer the neat, linear, step-by-step procedures of a quantitative paradigm but is rather characterised by a simultaneity of focus, with iterative, dynamic, interactive and recursive patterns of investigation. For this reason, Waite prefers the term, 'understandings' (1992a:354) for what is derived from the analysis of qualitative data. Nowhere is this more evident than in the first stage of analysis. Here, in the descriptive phase, data are collected and analysed at the same time. Accordingly, the researcher interviewed subjects (usually singly but in three instances, in pairs). Then, the log was created: first by listening to

[1] An alternative - verbatim transcripts - was avoided not only because it is time-consuming and expensive, but because this procedure had already featured prominently in the primary prong of the research.

the tape in its entirely; and then by listening closely, pausing regularly to key in the main points.

4.3.2.2 Unitising the data

The next step with each log was a process which Guba and Lincoln call 'unitising' (1983:314): creating the minimal, heuristic units of information that will later become the basis for determining conceptual categories[1]. Importantly, the units are generated from the data, not imported from elsewhere. Because 'it is the research problem that dictates the specifications of the research instrument (Fleet & Cambourne, 1989:8) it is unlikely that an imported (etic) system would 'fit' as snugly as a self-generating (emic) one. As a mechanism for quantifying qualitative data, the coding procedure allows a researcher to access the major themes in each interview. Accordingly, all information that could be itemised as a concern or issue or descriptor was recorded in the logs' margin notes, with an orientation, at this stage of the analysis, towards 'inclusiveness' rather than 'exclusiveness' (Guba & Lincoln, 1983:314). Each unit was given a code: 'a word or phrase that best describes an emergent theme identified in a block of text' (Baxter et al. 1992:216). At the same time, where readily identifiable, 'key words' (Baxter et al. 1992:219) were highlighted in the text, as the preliminary stage of creating clusters of representative associated words which adhere to particular codes. Thus, for example, very early in the analysis, the code 'morale' had key words like 'shattered', 'destroyed', 'dumped on', 'crumble', 'annihilate' as terms used by supervisors to express the fear of the impact of their words.

As the study inevitably moves from the descriptive towards the conceptual, the key word procedure helps as a validity check to counter the dangers of abstraction. Highlighting and retaining supervisors' key words preserves informants' views in the language they themselves used; and hence keeps the analysis well grounded in the data. While language exerts an inevitable mediating influence, maintaining the original key words is a way of minimising the intrusion of the researcher's subjective bias.

One of the fallacies of the genre of research description is that it may suggest a neatness of process and a clarity of purpose and thought that misrepresent actual realities (Littlejohn & Melouk, 1987). This early stage of analysis is very fluid and essentially evolutionary in nature.

---

[1] This is akin to the notion of a 'discrete unit of text' in Willms et al. (1990:408).

The final set of codes did not emerge in one sitting: it required successive 'waves' of combing, checking and refining. Throughout this phase, the aim was, following Guba & Lincoln (1983), to develop categories that had internal homogeneity and external heterogeneity, with a minimum number of unassignable items. With collection and analysis happening together, later interviews were inevitably influenced by early analyses. This fact, integral to qualitative research, refines and enriches the investigation.

4.3.2.3 Corroboration

The next step was to 'comb' the interview logs for corroboration of the seven *a priori* hypotheses that governed the design of the interview schedule. As each log was created and coded, it was checked for confirmation or otherwise of the hypotheses (as reported and displayed in Ch. 6). The process required listening again to the tape in its entirety and checking the log for fidelity to informant views. In particular it required attention to openness, that is, forcing oneself to hear what was said rather than what may have been sought. In this sense, effort was made to follow Hycner's recommendations for bracketing and phenomenological reduction: suspending researcher meanings and permitting the event to emerge as a gestalt: 'a conscious, effortful, opening of ourselves to the phenomenon *as a phenomenon* (1985:280, emphasis in original).

4.3.2.4 Saturation

The process of combined data collection and on-going analysis continued until the data were saturated (Glaser & Strauss, 1967). This point was determined following Guba and Lincoln (1983): when incoming data produces tiny increments of new information in comparison to the effort expended to get them; when regularities and a sense of 'integration' emerge; when the data suggests over-extension with new information failing to contribute usefully to the study. Closure of the data collection phase comes when the researcher, resourced by the wealth of the data bank, begins to believe in her own 'knowledgeability' (Glaser & Strauss, 1967:225). Given the richness and vastness of data elicited through interviews; the increasing familiarity of emergent patterns; the fact that the focus of the study is on qualitative not quantitative issues; and the awareness that collection could extend indefinitely if not monitored (Merriam, 1991), the decision was taken to cease collection when saturation - sensed from around the twentieth supervisor - occurred after the thirtieth informant.

The 30 interviews logs, now coded, with key words highlighted (where relevant), were now sufficiently substantial to be considered the 'primary resource package' (Patton, 1980, cited by Merriam, 1991:126). They were then organised chronologically into a data bank to enable ease of access and handling for the next phase, the intensive data analysis.

4.3.3 Intensive data analysis: 'mining'

The next stage required intensive 'mining' of the interview logs. The first step was to reduce the large number of descriptive categories to a smaller, more manageable number of conceptual categories (App. 27). These latter were obtained following the principle of Lincoln and Guba's convergent-divergent, card-sorting procedure (1985).

It was clear, following this step, that the conceptual categories fell into two types. The first of these is best described as minimal units representing emergent themes. Being minimal, they lent themselves readily to key-wording for they were essentially notions and concerns that were repeatedly expressed, albeit in a myriad of different ways, by the informant supervisors. There were six such 'cluster codes', detailed in depth in Ch. 6.

The second group were notions that were inherently more complex, and as such resisted the key word procedure. These were categories like 'building trust', 'cushioning bad news' and 'perceptions of conflict', where each category subsumed not inconsiderable sub-classes of information. These were established as 'bins', and the next step was to reduce them to a manageable and parsimonious number of high-priority issues. This necessary pruning was achieved by returning, as Goetz and LeCompte (1984) recommend, to the original questions that shaped the research investigation. As the final group of conceptual bins, along with the cluster codes, will become the basis of the final report, returning to the original research questions keeps the report focussed and reminds the researcher of the audiences for whom the study was originally intended. In this way, the categories were reduced to eight conceptual domains or 'bins'.

The researcher then returned to the data base contained in the primary resource file and proceeded to work through the 30 interview logs, mining them methodically for information that pertained to the bins. This involved fleshing out the categories and rendering them more robust by grounding them firmly in specific units of information gleaned from the data. Alongside the clusters of key words, these serve as empirically-grounded evidence of the

fundamental issues pertaining to supervisors in their role of deliverers of criticism. The report together with data displays makes up Ch. 6.

4.3.4 Research rigour

4.3.4.1 Validity

To be of value, research must inspire trust and credibility (McMillan & Schumacher, 1989; Merriam, 1991). Internal validity has to do with the question of how one's findings match reality; with the extent to which an investigation is actually measuring what it claims to be measuring (Nunan, 1992a:62).

A comment on the intrusiveness of method is warranted. In his attempt to refute the contention that qualitative and quantitative research paradigms are philosophically dichotomous Ratcliffe (1983) offers an illuminative perspective that aims to show that all research is essentially qualitative (c.f. Wolcott, 1988). Within Ratcliffe's philosophical frame of reference, three points of significance warrant mention. The first is that 'data do not speak for themselves'; they require an interpreter or translator; and this fact makes all knowledge 'fundamentally assumptive in nature' (1983:149). Secondly, the process of observing or measuring a phenomenon changes it; and this applies as much to 'hard' physics as much as it does to, say, interviewing. Thirdly, the various means by which reality is represented (numbers, equations, words) are all 'symbolic representations of reality' (1983:150), not reality itself. In the light of these observations, validity must be approached in terms of investigating the researcher's experience rather than in terms of 'assessing the isomorphism' between data collected and an external reality (Merriam, 1991:167) . Pure description, then, or in Wolcott's words, 'immaculate perception' (1988:18), is illusive.

4.3.4.1.1 Internal validity

There is general consensus that the major strength of qualitative research is its high internal validity, derived in particular from data collection and analysis techniques (Goetz & LeCompte, 1984; LeCompte & Goetz, 1982). In the present study, a lengthy collection/analysis period provided many 'opportunities for continual data analysis and comparison to refine constructs and to ensure the match between scientific categories and participant reality' (LeCompte & Goetz, 1982:43).

Secondly, using the interview as the means of data collection ensures that the language of inquiry is in alignment with the realities of subject informants: this personalistic contact with informants means the process of data collection 'is phrased more closely to the empirical categories of participants and is formed less abstractly than instruments used in other research designs' (LeCompte & Goetz, 1982:43). This step was reinforced by the 'key word' coding procedure, aimed at keeping the language of the report well grounded in the language of the informants. A third source of strength derives from the naturalistic context in which the data occurred: supervisors relate their own prior experience of supervision; there is no *a priori* control of events under discussion; the events and experiences called on are embedded in informants' own professional contexts. In qualitative research, one assumes that 'the meaningfulness of human actions depends on the contexts or situations in which these actions, feelings and perceptions occur' (McMillan & Schumacher, 1989:187). This research was conducted 'in natural settings that reflect the reality of the life experiences of participants more accurately than do contrived settings' (LeCompte & Goetz, 1982:43). Naturalness is defined by an absence of reactivity, artificiality and a minimum number of observer effects (Smith & Glass, 1987, cited by Wiersma, 1991:241).

Fourthly, there is 'researcher self-monitoring' (LeCompte & Goetz, 1982:43). The very nature of qualitative research - with its constant probing, self-conscious re-appraisals, trial and evaluation, and monitoring of bias and other threats to error - serves as a bulwark against threats to internal validity. This alertness of mind was most particularly experienced in the necessary tension, mentioned earlier, between, on the one hand, the corroborative nature of the investigation, pursued as part of the triangulation protocol, and, on the other hand, the need to be open and guard against perceptual bias. Through such disciplined subjectivity (Erickson, 1973), the ethnographer shores up validity through deductive processes (asking how has the data been influenced) as well as inductive processes (asking what sources of contamination may yet generate influence on the data).

Another source of internal validity is the exposure of the 'researcher's position' (Merriam, 1991:170) - a self-conscious endeavour through the entire process, but especially at the outset, to report frankly the researcher's 'assumptions, worldview and theoretical orientation' (1991:170). It is hoped that this quality is achieved in the present study, most particularly in the introductory chapter which examined professional and contextual issues; and through the statement of linguistic orientation in the conceptual framework that preambles Ch. 5 where the notions underpinning the pragmatic analysis of the supervisors' language are set out.

Lastly, the internal validity of the study was reinforced through other self-consciously implemented, self-monitoring procedures. Three of these were used: each in some form involved the 'recycling' of emerging constructs through informant member groups (Guba & Lincoln, 1983:319). One such was 'host verification' in which basic constructs that emerged from the data - the emergent conceptual categories - were re-played to a sample of 'stake-holding' members of the informant group[1]; and their responses ('member checks' - see App. 17) were elicited (Guba & Lincoln, 1983; Roberts, 1990; Zahorik, 1988). As Tabachnick points out, validity is strengthened if those involved agree that the report is 'accurate in its descriptions and fair in its interpretations' (1989:160). Guba and Lincoln place a very high value on this check of 'contextual appropriateness': 'who is in a better position to judge whether the categories appropriately reflect their issues and concerns than the people themselves?' (1983:97). In essence, this meant that the 'understandings' (Waite, 1992a: 354) of the researcher were checked against those of the stakeholding sample to check the degree of 'fittingness' (Guba & Lincoln, 1983:104). Their comments added further insights to the report.

A second recycling step that resulted in broadening the representative sample in the data base, operated through a grounded questionnaire (App. 18). The questionnaire is grounded in the categories that emerged from the mining of the interview logs of the 30 base-line informants. Forty respondents (different from the 30 of the base-line study) were asked to validate the issues by responding to a survey using Likert scales. Their responses were analysed to confirm or disconfirm emerging categories (Guba & Lincoln, 1983). These are discussed and displayed in Ch. 6.

The third appraising procedure was the use of a final check, called 'phenomenon recognition' (Guba & Lincoln, 1983:186). This involved sending a draft of Ch. 6 - the ethnographic report representing the investigator's 'reality' - to a sample of those who 'live it', and eliciting their comments on the adequacy with which it represents their lived experience (App. 19). This endeavour aims to establish 'the credibility of findings with relevant audiences' (Guba & Lincoln, 1983:115). As the six people to which this was sent are colleagues in the field, the measure also served as 'peer examination' of emergent themes (Merriam, 1991:169). Again, their comments were sewn into the final report. Such measures both reinforce 'the truth value'

[1] The 'stake-holders' came from the following contexts: adult EFL; adult ESL; secondary ESL; pre-service TESOL training; and in-service teacher development

of the report and support ethnography's goal of providing a rich, thick and resonant description of participants' meanings.

As a number of writers on research method have shown, the classic threats to the internal validity of quantitative research, as treated by Campbell and Stanley (1963), have to be re-interpreted in qualitative research, where they are defined and handled differently (Guba & Lincoln, 1983; Lecompte & Goetz, 1982; McMillan & Schumacher, 1989). The major threat is that posed by human instrumentation. This weakness is off-set by the peculiar advantages wrought by that same humanity: 'flexibility, insight and ability to build on tacit knowledge' (Guba & Lincoln, 1983:113) as well as empathy. In addition, the researcher is by no means helpless, able indeed to take steps to 'shore up' internal validity (Guba & Lincoln, 1983:115).

In regard to threats related to human instrumentation, one counter is to guard against observer effects: through spending adequate time in the field; monitoring for halo effects (Cook & Campbell, 1979) and contamination (Borg & Gall, 1989); and avoiding over-reliance on a few key informants. As well, researcher bias is guarded against by the conscious creation of the dual identity of outsider-insider. This deliberate duality avoids both the danger of 'going native' (LeCompte & Goetz, 1982:47) and the dangers of being too removed from the data source. The dual persona, while difficult to implement, is considered an ideal balance permitting in the end an 'authentic presentation of the participant world' (LeCompte & Goetz, 1982:47).

Another means of avoiding observer effects is to increase the breadth of data through the number of informants and the range of contexts tapped. The 30 supervisors of the base-line study draw on experiences from as many different sites which display a great range of diversity in setting, context and purpose. (The data description in the early part of Ch. 6 provides more details of this). Corroboration across such a wide range of situations lends robustness to the study.

4.3.4.1.2 External validity

External validity relates to the extent to which the research outcomes are able to be extended to other groups; and is jeopardised by effects that threaten its comparability. This notion of applicability is recast by Guba and Lincoln (1983:119) for naturalistic contexts as 'fittingness' - or the degree of fit between the research context from which the outcomes were derived and the context to which they are to be applied. Guba and Lincoln distinguish between the

quantitative term 'generalisations' ('bare-bones statements stripped of all contextual reference') and the qualitative term 'working hypotheses', which suggests a strong contextual relationship (1983:119). The notion of 'working hypotheses' in place of generalisations was proposed by Cronbach (1987). His view is shared by Patton, who argues that qualitative research seeks perspectives rather than truths through assessments of 'local decision-makers' theories of action' (1980:283), not quests for verification of universal theories. The advantage of the 'thick description' (Geertz, 1975:6) that characterises the ethnographic report is that it provides the details of context that enable an estimation of fittingness to be made with another context. Comparison is based on the establishment of 'the typicality of a phenomenon' (LeCompte & Goetz, 1982:51). It is hoped that the thorough description of the research components (in this chapter) and the detailed report of the investigation (Ch. 6) together provide a basis upon which 'comparability' and 'translatability' (McMillan & Schumacher, 1989:194), or what Bassey calls 'relatability' (1981:85), are afforded. The 'burden of synthesis' (Guba & Lincoln, 1983:117) - that is, determining the extent to which the study's outcomes apply to another situation - then rightfully shifts to the recipient of the report - e.g. an interested practitioner within another yet-to-be-researched field of interest (Walker, 1980).

External validity is strengthened by multi-site studies and cross-site corroboration: 'if a phenomenon seems to be consistent across a number of sites, its generalisability is increased' (Wiersma, 1991:242). As described above, the interviews that form the data bank of this study have in common their TESOL educational connection, but beyond this, they derive from a wide range of sites and contexts. Such variation shores up the claim to external validity.

4.3.4.2 Reliability

Reliability is about replicability, or the extent to which one's findings can be reproduced. It is a problematic notion in the areas of social sciences, such as educational research, because of the human nature of what is being studied: complex, multi-faceted, highly contextual, largely invisible, dynamic and fluid. Qualitative research does not aim to isolate fixed and inviolable laws about human behaviour, but rather 'seeks to describe and explain the world as those in the world interpret it' (Merriam, 1991:170). The processes are essentially 'personalistic', (McMillan & Schumacher, 1989:188) with no two investigators, observing, interviewing, and analysing in exactly the same way. An onion has many layers; and who is to decide which layer will be studied? (Guba & Lincoln, 1983): replication within human

studies is largely impossible because of the fluid nature of what it means to be human (1983:116):

> Chinese mythology tells of the wanderer who asked a philosopher whether it is possible to cross the same river twice. The philosopher replied that it is not possible to cross the same river even once, since the river is flowing by and changing as the crossing is taking place.

The extreme view is that reliability and internal validity are inextricably interwoven; and because a 'demonstration of internal validity amounts to a simultaneous demonstration of reliability' (Guba & Lincoln, 1983:120), some writers recommend side-stepping issues of reliability in favour of 'shoring up' internal validity (Guba & Lincoln, 1983).

4.3.4.2.1 Internal reliability

Internal reliability refers to the extent to which an independent investigator, analysing the primary data, will come to the same conclusions as the original researcher. Arguing against replicability in the tradition of 'objective' scientific research, Hycner (1985) claims that the nature of qualitative research dictates that differences among researchers are inevitable. The real issue, he claims, citing Giorgi (1975), is 'whether a reader, adopting the same viewpoint as articulated by the researcher, can also see what the researcher saw, whether or not he [sic] agrees with it' (1985:298). The findings must be consistent, dependable and make sense in the light of the data collected (Lincoln & Guba, 1985). In order to come to such an assessment, an independent investigator must be able to follow the original researcher's 'decision trail' (Guba & Lincoln, 1983:122). Accordingly, Guba and Lincoln translate the term 'reliability' into 'auditability' for use in naturalistic contexts (1983:104). The report must be able to serve others as an operating manual so that they might access the decision-making processes of the original researcher. In this study, the audit trail for the ethnographic study is found in Chs 4 and 6.

Apart from an accessible trail, internal reliability is shored up through other means, described in detail by LeCompte and Goetz (1982). The first of these is the attempt to keep descriptors as low-inference as possible. Thus in the first layer of analysis, the coding of blocks of texts in the interview logs, the codes are phrased very precisely and concretely. These later become the basis of the higher-inference conceptual categories, derived through the sorting process described by Lincoln and Guba (1985). In the fleshing out of categories and 'mining' of logs for the more complex 'bins', attempts are made to maintain authenticity by keeping

descriptions as close as possible to the original language of the informants. In addition, as described earlier, colleagues in the field are called on as 'participant researchers' to review, corroborate and validate findings through 'phenomenon recognition'. The use of mechanically-recorded data allows access to primary data and confirmation of veracity; and low-inference descriptive logs seek to be faithful (Giorgi, 1971), relatively instrument-proof, one-step-removed records of interactions.

4.3.4.2.2 External reliability

External reliability refers to the replication of the original study: could another researcher working with different data and constructs match the result? As with internal reliability, replicability is not the issue in qualitative research that it has traditionally been in quantitative studies. Many of the safeguards have already been discussed. The burden of replicability comes down, in the final analysis, to issues of 'care' and 'explicitness' (Nunan, 1992a:62, drawing on LeCompte and Goetz, 1982): care taken in the collection and analysis of one's data; and explicitness in the way in which these are described in the audit trail. The descriptions offered on the subjects, settings, contexts and conditions under which the data were elicited (Chs 4 & 6), as well as the detailed outline and display of analytical methods (Ch. 4), have sought to be both careful and explicit. It is the researcher's duty to provide these: the responsibility for comparability and transferability is passed on to the interested researcher who wishes to see whether for example, in educational contexts other than TESOL (such as other subject disciplines), or perhaps fields outside of education (such as management), the findings pertaining to critical feedback have wider applicability.

Overview

This concludes the research method in the ethnographic study. The reader may wish to go straight to the report of findings (Ch. 6); or alternatively, following the chronology of the thesis, move on now to Ch. 5, embracing the linguistic findings.

4.4 The third prong: research method in the controlled experiment

The third arm of the triangulation protocol was achieved through a controlled experiment. The spotlight, heretofore focussed on the supervisor, now shifted to the supervisee. There was a

switch, too, in research method from qualitative-interpretive to a quantitative-psychometric. The aim of the experiment was to measure subject responses to controlled input - variously mitigated supervisory language - under simulated conditions. The data yielded were processed by statistical aggregation using SPSS software.

While this chapter has been dedicated to research method in the two other prongs of the study, the third prong is integrated with the findings and discussion of the experiment (Ch. 7). This follows the tradition of psychometric experimental reports within the empirical-analytical paradigm, where the discussion of research method is placed alongside the description of data and discussion of findings. The reader is therefore referred to Ch. 7.