Performance features in clinical skills assessment

Linguistic and cultural factors in the Membership of the Royal College of General Practitioners examination

Performance features in clinical skills assessment

Linguistic and cultural factors in the Membership of the Royal College of General Practitioners examination

Celia Roberts, King's College London

Sarah Atkins, University of Nottingham

Kamila Hawthorne, Cardiff University

Acknowledgements

What are Knowledge Transfer Partnerships?

KTPs help UK organisations to improve performance through developing collaborative partnerships with academic institutions. Furthermore, KTPs aim to foster continued collaborative relationships, as has been the case with this research project. Sarah Atkins, as an ESRC Research Fellow at the University of Nottingham, is continuing to work with the RCGP and this includes the development of CSA training materials.

How to read this document...

This is an unusual document and readers may find this brief guide useful in navigating their way through it.

This study came out of a 'Knowledge Transfer Partnership' and a two-year period of close cooperation between university experts in linguistics and members of the Royal College of General Practitioners. The report, therefore, inevitably crosses different disciplines and uses an amalgamation of terminology – from linguistics, pedagogy and medical communication. Where possible, we hope to have presented short explanations of terms that may be new to the reader, as well as summaries of key findings.

Key:

Throughout the document, you'll find text highlighted in coloured boxes.

Summaries – For the reader in a hurry:

Given the amount of analysis we covered over two years, this is a long report that attempts to explain all that we found. However, for those with less time to read the analysis itself, summaries are given in green boxes at the start of each chapter. Following these summary boxes will give an overall picture of the findings and position of the research.

Definitions – For new terms and ideas:

Where a term from linguistics has been used, a definition appears in a yellow box that can be referred to for an explanation. This was not an arbitrary decision. One of the declared objectives for the project was to develop a new analytic vocabulary for the CSA, and these 'Linguistic terms' boxes begin to establish that. Some aspects are developed further in e-learning materials.

Examples – Direct quotes from CSA interactions:

A central tenet of Applied Linguistic studies is to closely examine what is actually said in 'real-life' interaction, rather than just intuitively what we think is said. Wherever a close analysis is made of the data, the quote appears in blue. 'CAN:' refers to the candidate's line and 'RPL:' to the role-player's. Role-player is usually shortened to 'RP' in this document.

The report, therefore, presents a comprehensive account of two years' worth of analysis of different but interlinked data, under the overarching question of how linguistic and cultural factors may play a role in candidates' success or failure in the CSA. Chapters 1-2 describe the overall principles and method, Chapters 3-4 present results from a linguistic analysis of spoken data, Chapter 5 a thematic analysis of examiner video feedback style discussions and Chapters 6-9 an argument for the implications of our findings.

Further work and analysis can be found in the Appendices, which are referred to throughout the main document, but these do not have to be read to understand the core findings. The appendices are available on line:

http://www.kcl.ac.uk/sspp/departments/education/research/ldc/publications/index.aspx

Executive Summary

Introduction

This independent research project was a Knowledge Transfer Partnership between King's College London and the Royal College of General Practitioners, funded by the Technology Strategy Board and the Academy of Royal Medical Colleges (2011–13). The aims of this research project were to:

1. understand the performance features of the Clinical Skills Assessment (CSA), part of the licensing exam for all GPs in the UK, and, in particular, to investigate the extent to which linguistic/cultural factors contribute to poor performance.

2. develop an analytic framework and design associated materials to raise awareness among examiners, GP trainers and candidates of the linguistic and cultural demands of the exam.

The stimulus to the research was the recognition that international medical graduates and black and minority ethnic UK-trained graduates were less likely to pass than white UK-trained graduates. For the purpose of this study all graduates who trained abroad both from the EU and elsewhere are included in the category International Medical Graduates.

In order to understand performance, it is necessary to understand the context within which such performances come to be produced. So, this research investigated not only how candidates actually behave in the CSA but also how the specific conditions of the exam operate to determine this behaviour. These conditions include the fact that the assessment is of simulated consultations and that the CSA consists substantially of talk and interaction within a patient-centred model. In addition to candidate performance in selected cases, therefore, role-player (RP) behaviour in those cases and examiner feedback on video-recorded segments of CSA cases were also analysed together with aspects of the CSA paperwork.

Underpinning this research is a central paradox in institutional life: how to be fair and maintain standardised and universal criteria in an increasingly diverse society. This is a paradox that all institutions who assess and select have to face. While this research only covers the CSA, many of the issues will be equally relevant for any simulated clinical consultation and OSCE type exam (e.g. Wass et al 2003, O'Grady 2011, De la Croix and Skelton 2009). The conclusions and implications, therefore, must be read in the light of their more general significance for exams of this genre, in both undergraduate and postgraduate contexts, particularly with the increasing globalisation and movement of medical personnel between countries in the 20th and 21st centuries.

Executive Summary

Research methods

Our research used qualitative and quantitative sociolinguistic methods, supported by ethnographic information (see further 'A note on methodology' below): (i) We video recorded 198 consented candidates over two exam diets in February/March and May 2011. We made a detailed analysis of 40 cases across the dataset, as well as viewing all tagged 'diversity cases' from within this sample (the full case bank amounts to over 650 cases). The 40 cases excluded examples of candidate performance where there were clear and important clinical errors/mismanagement. Cases were selected from the full range of 198 cases video recorded for the research and represented a mix of routine and complex cases, both cases with a predominately medical focus and cases with a predominately psycho-social focus.

Complex psycho-social and emotionally demanding cases made up just under a fifth of the dataset. Pass, fail and borderline cases were reflected in this sample together with UK trained candidates and those trained abroad These cases were subjected to both a broad based interactional analysis as well as micro-analysis at the level of individual speaker turns. The advantages and limitations of this methodology are discussed in more detail in the main report

(ii) We reviewed the CSA process paperwork (including case materials and marking and feedback schedules) to identify how the three assessment domains were used. (iii) We analysed examiner feedback, with current MRCGP CSA examiners, on a range of video clips both from the data set of 40 and from the whole 198 case sample.

This partnership between King's College London and the RCGP, along with researchers from Cardiff University and, more recently, the University of Nottingham, is the first of its kind and, together with the RCGP's publication of examination data in Annual Reports and published papers, is an indication of the RCGP's willingness to be open and transparent in its intentions to conduct a fair but rigorous clinical examination.

Candidate Performance

(from analysis of the 40 video cases)

Performance was analysed at two complementary levels: a more broad-based level and an in-depth detailed (micro) level.

Broad based analysis

We made several overview analyses of our data, looking at the structure of the cases and the lexical content of talk (see 'a note on methodology' below). These broad-level analyses did not show important differences between candidates who performed well and those who performed poorly.

- Candidates structured and paced their consultations in broadly similar ways.

- There were no obvious differences in the amount successful and unsuccessful candidates talked and they used similar medical and social language, including the typical phrases of the CSA e.g. 'I understand how you feel', 'OK, all right', Can you tell me a bit more about..?'.

- There were, however, two exceptions to this general finding: (i) poorer performing candidates tended to show late data gathering behaviour (ii) weaker candidates were more likely to be interrupted by role-player (RP) patients seeking clarification, when giving explanations.

- However, at the micro-analytic level differences were clear.

Micro analysis

A qualitative, micro-analysis of the 40 cases reveals that poorly performing candidates were more likely to:

- Have difficulties giving explanations to RP-patients.

- Have misunderstandings in the consultation.

- Have more difficulty repairing misunderstandings.

- Experience more moments of misalignment with the patient that would impact on the unfolding consultation.

- Sound formulaic to examiners.

Often, reasons for failure could not be identified in any one particular moment of the consultation, but rather represented an accumulation of these small, micro-level difficulties in communication.

Candidates performing well were more likely to:

- Align with patients but also balance this with institutional requirements.

- Make the talk more conversational.

- Combine clear explanations with alignment to patients.

- Use typical CSA phrases, while finding ways of customising them to avoid sounding formulaic.

- Communicate with fewer hesitations and false starts, although all candidates did experience some of these.

- Identify and repair misunderstandings and misalignments more easily.

- Convey their stance and intentions to the examiner, by commenting out loud on what they were doing – 'metacommunication'.

Executive Summary

Role-players

- RPs played a powerful role in the shaping of the consultation, which was a function of their role.

- We found no evidence of language-focussed discrimination in the consultations in the micro analysis of the discourse.

- Although RPs could sometimes, 'save' or 'sink' candidates at moments of the interaction, this did not happen with any consistent pattern, according to the background of the candidates or their score for the case.

- RPs tended to be the ones who highlighted misunderstandings in the consultation and these moments were found to occur more frequently than is evidenced in 'real-life' patient consultations (when compared with other corpus databanks of real Dr/patient consultations). But this is a factor of assessment design and the nature of their role, rather than intentionally difficult behaviour by the RP.

Examiner feedback

(from examiner video feedback discussions of video clips)

- Examiners understood, and were sympathetic towards the stress and issues facing candidates, particularly international medical graduates (IMGs).

- Examiners, trained to be very alert to cues in talk and interaction, made rapid and important judgements of manner and affect. These related, largely, to how candidates sounded. How someone sounds is usually described, in society at large, in terms of accent i.e. a speaker's pronunciation. This overlooks the importance of other features of talk such as intonation, pace and rhythm. There were no comments about candidates' pronunciation that indicated that they were considered hard to understand. However, many of the comments examiners made in response to taped CSA clips related to intonation, pace and rhythm which are, generally, subliminal aspects of talk but are crucial in evaluating interpersonal skills and coherence.

- Although there were individual differences in how candidates were judged, there was broad agreement in the areas examiners chose to comment on i.e. largely on manner and affect. Such agreement is not surprising since examiners are a relatively homogeneous group in terms of shared notions of 'common sense' in consulting, professional shared meanings and in marker training for the CSA.

- The descriptions of manner and affect used general, informal, evaluative language. There is a strong case for introducing an analytic language which has explanatory power for registrars, trainers and examiners.

Exam Design

Case design:

- The standardised framework of an OSCE type assessment necessarily gives a homogeneity to the cases which appeared to value a particular, articulate style of professional talk that occurs within a defined range of British consultations. Some types of consultation are difficult to simulate, for example, a 'patient' with several problems to discuss, a 'patient' who speaks little English, or where more than one person is attending the consultation.

- A cohort of cases are tagged as containing an element of 'diversity' – as defined by the Equality Act. All such cases, within the total 198 video recorded for this research, were identified and viewed by the research team. The tagged 'diversity' cases included a number of different culturally sensitive issues in a clinical context, but were conducted in English with RPs who used a local British interactional style and did not include any RPs from different linguistic backgrounds or culturally-specific styles of communicating.

The role of talk:

- Candidates talk relatively more in the CSA than GPs in real consultations. The CSA is also more decontextualised in that it focuses the encounter on talk and interaction (for example, there is no interaction with a computer – one of the very different contextual conditions that changes talk and the focus of much interaction in everyday GP encounters).

- Small differences in talk in interaction can lead cumulatively to large consequences (supporting previous research e.g. Gumperz 1982) e.g. to judgements of affect and manner (see below). These small differences, as in many assessed settings of this type, can be amplified because of the intense focus on how talk is delivered and the interaction progressed.

- The CSA has a particular linguistic 'fingerprint' that involves talking slightly differently from much 'everyday' UK General Practice and requiring additional layers of communication i.e. managing simulations, dealing with explicit interventions from role players and displaying knowledge to the examiner.

- Suggesting to candidates that they 'practise as you would with your own patients' is necessary, but not sufficient advice.

Executive Summary

Interpersonal skills domain:

- In some instances, the exam is trying to assess areas that our evidence shows may not be best assessed in a highly standardised and simulated exam.

- The interpersonal skills domain (IPS) is particularly problematic. This is an assessment issue identified both from prior pedagogic research (Howie et al 2004, Stepien & Baernstein 2006) and from the findings of this project. While communication is a vital component of GP consulting, it is not possible to objectively judge some aspects of IPS i.e. empathy, rapport or sincerity from the outside. Empathy is an inner state, experienced (or not) only by someone to whom it is directed. On the basis of our analysis of a highly standardised simulated setting, examiners can assess whether candidates use 'empathic' expressions but not whether patients experience empathy. This could be done in workplace based assessment with real patients, and is important to test for, but cannot be reliably tested in the CSA. However, there are other, more objective aspects of communication that can be tested in the CSA, for example, the degree of role-player/candidate alignment and giving coherent explanations.

- Interpersonal effectiveness depends on manner and affect, aspects of which are realised through the largely unnoticed features of intonation, rhythm and pace. These hidden aspects are the most difficult features to be trained in for candidates, and for trainers and examiners to analyse, thus making IPS a difficult area to assess for everyone, and making it a problem area when considering differential pass rates.

- Our analysis of the case specific marking schedules found that the IPS is often assessed outside its domain as well as within it. While the 3 domains of the CSA are not orthogonal (i.e. the domains overlap), IPS case specific marking 'leaks' into the data gathering (DG) and clinical management (CM) domains. Leakage the other way is infrequent.

- As well as the formal criteria, evaluation of manner and affect are inevitably made at all points in the consultation, as evidenced from examiner feedback discussions when watching a full range of cases from the 198 videos. This indicates a likely imbalance across the three assessed domains.

- So Interpersonal Skills are both explicitly and implicitly assessed, giving an intense focus on this linguistically and culturally demanding aspect of the exam.

Implications of these findings

Implications for candidates

All failing candidates:

- Some of these implications relate to all poorly performing candidates, others largely to those who are international medical graduates (IMG) (see below).

- There are identified features of consulting skills in poorly performing candidates, listed above, which suggest the need for focused training, support and preparation in these areas.

- It is a linguistically/culturally demanding, 'talk heavy' exam that requires considerable communicative fluency. Our data do not suggest that this is an issue for British ethnic minority UK trained candidates but it has a particular bearing on IMG candidates.

The effect of the exam on IMG candidates:

1. The relatively decontextualised nature of the exam makes it a 'talk heavy' assessment from which several effects flow:

   - There are communicative performance factors that contribute to the gap in success rates between IMG and other candidates. These result both from aspects of how IMG candidates talk and interact with RPs and from the unintentional effects of the exam design described above.

   - The higher rate of failure of this group (aside from clinical errors and mismanagement) relates to some lack of clarity in all types of explanations to RP patients, higher rates of misunderstandings with RP patients, and examiner perceptions of the manner of candidates – how they sounded, e.g. formulaic, not engaging with RP patients.

   - The additional communicative demands of coping with simulation and its unfamiliar complexities appear to fall heavily on this group e.g. having to imagine and talk about aspects of the case or the RP patient's background which are not in the case documentation given to candidates. These additional demands are evidenced by poor performance and comments by examiners of 'formulaic' and 'clunky' consulting.

   - While there was no evidence that differences in pronunciation cause misunderstandings for either candidate or RP, the rhythm and intonation of speech which convey both information and attitudes (and so much of a candidate's 'manner') can vary between candidates and affect how performance is rated.

Executive Summary

2.	The additional emphasis on assessment of interpersonal effectiveness in the exam (as found in the leakage of the IPS domain across all three domains and the focus on manner and affect in examiner feedback) unintentionally adds to the weight to the exam. Interpersonal skills are the most culturally-specific and linguistically demanding aspect of the exam for this group and the skills most based on examiner instinct. The widely recognised (if implicit) patient-centred model currently favoured in UK practice can produce particular demands on candidate manner and communications. Only successful candidates customise the model to avoid sounding formulaic, insincere, or jarring and this customising of CSA patient-centred phrases is an additional challenge for many IMG candidates.

3.	The case design and use of RPs is based on a local British interactional style and does not assess or incorporate the multilingual expertise of IMG candidates, although society is increasingly diverse and one third of candidates are IMG.

4.	There is an accumulative effect of all these performance requirements and the design of the exam on the actual performance of individual candidates.

Educational interventions: the context

•	These materials focus on developing communication in those areas which the research has identified can lead to lower marks in the exam and which the CSA can most readily assess. Other aspects of the consultation can be assessed in the work environment as part of work based assessment.

•	The sociolinguistic performance features which contribute to lower exam ratings are difficult to identify and teach. In many respects, they constitute the hidden curriculum of the exam. There is the danger that if these features are made explicit and translated into teachable items they may reproduce the very problem that they are attempting to solve e.g. make candidates sound more trained and less sincere and sensitive. The materials are designed to avoid this by basing the learning on awareness raising of candidates' own performance.

Educational interventions: the materials

•	We are preparing e-learning modules, 'Inside the CSA', so that, by using real video recordings from the CSA to develop a new analytic language, registrars can learn to monitor their own performance in consultations (both real and simulated), address the features of poorly performing candidates outlined above and work on the features of successful candidates. This new analytical language will help registrars, trainers and examiners to: (i) shift away from assessing consulting skills in terms of inner psychological states (such as 'empathy', which is hard for some to simulate and can only be experienced by those to whom it is addressed) towards a more objective and

interactional concept 'alignment' which focuses on mutual understanding (ii) learn new terms which encourage aspects of talk and interaction to be noticed more (iii) develop a more nuanced understanding of language varieties and how they are judged.

- Candidates can also learn and practise the elements of giving a good explanation, aligning with patients' agendas, and noticing and repairing misunderstandings in consultations.

- These new e-learning materials will be designed to be useful for both registrars and GP trainers. The messages from these materials and from this report will also benefit examiners and those going on examiner training.

- A new RCGP book on the CSA, co-authored by Alexandra Rolfe and the KTP team, will be published which is to be partly based on this research and will link to the e-learning materials.

Implications for the design of the CSA

The IPS domain:

Considerations for the Assessment Team:

- The IPS domain can be helpfully reconceived, replaced by a set of criteria based on the new analytic language, relating to communication in a professional encounter that can be assessed more objectively in the exam.

- Clarity, mutual negotiation of understanding and alignment (keeping communication channels open) can be judged. This would be a shift away from imagining the internal reactions of RPs and candidates and towards an assessment of communicative behaviour.

- For examiners, the focus should be on assessing how clear communication is and on a reasonable level of alignment to get the work of general practice consulting done.

Issues of diversity:

- This shift towards a less intense assessment of IPS could be done in a number of ways, including altering the IPS domain and how it is marked.

- Aspects of the consultation such as longer term relationship building and managing the interaction between patient, computer and doctor are best summatively assessed as part of the workplace based assessment.

- Analysis of diversity cases showed that multilingual communication (which forms an integral part of candidates' consulting experience in large urban settings) is not

addressed in the CSA. This is an area where IMG candidates are likely to have particular strengths and recognition needs to be given to their skills.

- In addition, working in challenging consultations across language barriers is a competence that should be assessed for all candidates since the CSA licenses GPs to work throughout the UK which is becoming an increasingly diverse society.

- It is almost impossible to simulate such consultations, but they are important and need to be included and given proper recognition within the curriculum assessment.

- Rather than talk of 'cultural bias' or not, there needs to be a debate about tolerances and communicative flexibility, about what are acceptable competencies in an increasingly diverse society and how, within these competencies, talk and interaction can be more explicitly addressed. 'Cultural bias' implies that there is a goal of neutrality that must be reached and that there is one 'culture', one way of doing things.

A note on methodology

This ethnographically-informed sociolinguistic project used a mixed method approach for understanding interaction and oral performance, drawing on both quantitative and qualitative approaches at two levels: a relatively broad-based (more macro) level looking at patterns across the whole data set and an in-depth, detailed (micro) level. There is no simple equation between macro and quantitative and micro and qualitative. In this research, quantitative and qualitative methods are used at both levels. These approaches and levels are complementary and necessary to achieve the goals of the project. They rely on the principle of looking closely at 'real-life' language data, collected and transcribed from fieldwork, in order to make any claims about linguistic behaviour.

Putting a magnifying glass on the detailed processes of interaction allows hidden aspects of the CSA to be made visible. Micro-analysis, therefore, which forms the bulk of the analytic work, provided some of our most useful findings. However more broad-based analysis was important in identifying interactional patterns which needed to be researched in more depth and in establishing the context for our analysis at the micro level. This broad-based analysis, resulting from coding and quantitative analysis of transcripts, identified certain performance features of candidates and examiner reactions. This analysis, while telling about the nature of the exam, is somewhat limited because of the relatively small database.

The unit of analysis at the micro-level is much smaller (i.e. at the level of word, phrase and sequence of interactive turns) and so provides a larger data base at this level for both qualitative and quantitative research. It is used to collect and unpack data in very different ways from traditional methodologies and so reaches levels of analysis that these other methodologies cannot reach.

Firstly, this small unit of analysis means that relatively large numbers can be used for

computational micro-analysis. This is a method called 'corpus linguistics' which takes the unit of the word as the basis of calculations to find patterns. The 40 recorded consultations from this specialised context gives a database of 85,000 words, from which clear patterns can be identified and statistically evaluated against larger datasets. The generalisability of this method comes from the fact that, within specialist datasets such as this, 85,000 words begin to exhibit linguistic patterning that is sufficiently repetitive to make statistical claims about the way talk is characterised in this setting and to find clear distinctions from general language use.

Secondly, qualitative micro-analysis uses an aesthetic of smallness and slowness to:

1. understand how talk, moment-by-moment, is produced and determined by a particular context and, especially in high stakes assessment, how small differences or discrepancies themselves produce a new context by which speakers may be judged.

2. identify the conditions for talk, even in a single telling moment, such as a misunderstanding, which are generalisable across broadly similar settings (such as OSCE type settings). For example, our analysis explains why, although all candidates use 'empathic' phrases, poorer performing candidates are seen as formulaic and so less interpersonally effective because the conditions of such an examination require a demonstration of 'empathy' spoken in a particular style.

The mixed method approach used in this research is open to challenges. The video data set is relatively small for corpus linguistics as is the examiner feedback data for thematic analysis. By contrast, 40 cases is a large data base for micro-analysis and only certain features, identified in previous intercultural research, could be analysed in depth.

What gives the findings the potential for more generalisable comment and application to other contexts, therefore, is not how far they can be said to be statistically representative of all candidates, but how the interactions themselves uncover conditions for talk that are common to many such settings. So, qualitative micro analysis: (i) helps to explain and offer solutions to interactionally based social problems (ii) sets detailed analysis of data within the specific conditions of their production so that previously hidden relationships and phenomena become apparent (iii) allows for applicability of concepts, patterns and phenomena to other contexts, provided these contexts meet the same conditions as those in the original data (iv) stimulates creative thinking and can disturb widely-held assumptions.

Abbreviations

Transcripts Key:

`CAN:`	– Candidate's talk
`RPL:`	– Role player's talk
`EXM:`	– Examiner's talk
`(.)`	– Very short beat pauses, less than 0.3 of a second
`(0.5)`	– Timed pauses (timed to a tenth of a second)
`pa-`	– Syllable cut off, false start
↑	– Rising pitch
↓	– Falling pitch
hhh	– Inhalation
≋	– Continuation marker (speaker continues across turns, no break)
≈	– Latching (one speaker commences immediately as another finishes, with no gap)

Brackets:

`[]`	– Overlapping speech. For example:
`⌊⌋`	

```
ACT:    runs in families  [so yeah yeah]
CAN:                      ⌊mmm mmm⌋
```

??	– Inaudible/unsure of transcription
??**xxx**??	– Audible word is articulated, but unsure of transcription

Contents

List of 'Linguistic Terms'

List of Figures

Contents

Appendices referred to are available at

http://www.kcl.ac.uk/sspp/departments/education/research/ldc/publications/index.aspx

Chapter One. Introduction

1.1 Research framework and background to the study

Identifying and understanding face-to-face performance in high stakes encounters is a difficult and contested task, particularly when outcomes related to racial/ethnic background are documented. This is because performance is made up of talk and interaction which are shaped by language and cultural practices. It is widely assumed that since we all use language, we know how it functions and are able to judge its effectiveness. In addition, in any debate on fairness, 'culture and language' are seen as standing apart from 'fairness' rather than being part of it.

In order to provide a helpful frame for thinking about these matters, the detailed data analysis and interpretation are prefaced by three overarching themes, expressed as metaphors, which relate to: (1) candidate performance and the challenges of the exam; (2) the process of assessing candidates; (3) and some reasons why there are such polarised views on the exam and its fairness.

1.2 'The fish out of water'

'The fish in water does not feel the weight of the water' (Bourdieu and Wacqant 1992)

This metaphor was developed to clarify the process whereby class differences and disadvantages are perpetuated by, in particular, the education system. It is now used more widely to look at group disadvantage more generally. It is based on the concept of 'habitus' – our ways of thinking, acting and interacting, our dispositions, which are the result of our personal history (Bourdieu 1977 93–4).

Our habitus fits well (or not) into particular fields e.g. formal institutions such as education or medicine, or, by contrast, informal settings such as being a member of the local hunt or football club. When there is a fit between habitus and field, we can act largely unconsciously – we do not feel the weight of the water – and we have what seems to be a 'natural' affinity with that field. This is what Bourdieu calls 'a feel for the game' (Bourdieu 1990: 61). When personal history and experiences do not provide such a fit, then an individual can feel a like a fish out of water. We all experience, at different times, the feeling of being fish out of water, feeling its weight, but luckily, in most circumstances, the evaluation of our performance does not have lasting consequences. The idea of individual habitus has also been expanded to 'institutional habitus' (Reay, David and Ball 2005) where the habits and practices of groups are embedded in the culture of an institution. So, like individuals, groups who work in institutions can experience its way of working as second nature.

When scrutinising the performance and assessment of candidates in the CSA, this metaphor helps in thinking about the experience of the weight of the exam on certain groups of candidates and to account for how the majority swim through the water more easily than others. The three year training period may make all candidates feel that they are fish in the GP water and yet the outcomes of the exam show that it is the IMG groups which, disproportionately, experience its weight.

Introduction

1.3 The magnifying glass effect

Talk and interaction are used by listeners to make assessments about speakers' origins, experiences and competences and how different/similar they are to listeners. Aspects of someone's identity and competence leak out through small features of talk and interactional smoothness or turbulence e.g. 'typical northerner', 'a caring manner'. In most settings, these may be registered (consciously or unconsciously) but are insignificant since the purpose of the interaction overrides these assessments. For example, the football manager from Spain in a television interview is judged on the quality of his opinions, not his intonation. In gatekeeping encounters such as job interviews or oral exams, the focus on oral performance amplifies similarities or differences. The talk and interaction are put under a magnifying glass so that small interactional moments can have large consequences. These may be one-off moments or be the cumulative effect of many. As E.M. Forster says in *A Passage to India:* 'A pause in the wrong place, an intonation misunderstood, and a whole conversation went awry.' (Forster 1936: 267). This metaphor of the magnifying glass helps us to understand the process of assessing oral performance and how some aspects of it are amplified in this particular setting. It also helps to account for, paradoxically, both differences of opinion between and yet broad agreement amongst those who share an institutional habitus when assessing individuals' performance, as Chapter 5 shows.

1.4 'Hidden in plain sight'

'The aspects of things that are most important for us are hidden because of their simplicity and familiarity. (One is unable to notice something – because it is always before one's eyes.) … And this means: we fail to be struck by what, once seen, is most striking and most powerful.'

[Wittgenstein, *Philosophical Investigations*, 1953, No. 125]

Talk and interaction are the fabric of the CSA encounters. Physical examinations take up only a small amount of time. Since talk and interaction are what we do all the time, how they actually function and are processed is largely 'hidden' and yet we continuously judge people on the basis of them. So, in issues of 'fairness' or 'bias', talk and interaction are obvious, and yet 'hidden' as factors when scrutinising the design, processes and outcomes of institutional assessment. For example, in competence based job interviews, telling a well-structured story is essential to success but this feature is hidden in plain sight (Roberts and Campbell 2006). This phenomenon helps to explain the gap between the widely accepted quality of the CSA, on the one hand, and the critique of it as 'biased', on the other, since the hidden aspects of the exam are not easy to identify by either side. Both sides experience the exam – it is in plain sight – but the evidence for understanding the gap between the two positions is hidden in taken for granted ways of acting and interacting.

Questions of fairness and equality become more charged when 'race' and 'culture' are added to the mix. The issues of how IMGs come to fare so much less well than other groups are, of course, complex. At the most general level, the current CSA practice and its outcomes raise the

big question of how institutions can offer standard opportunities and make objective and fair decisions in conditions of superdiversity (such as the multicultural population of the UK). What are the norms in such societies? What makes a good GP in these conditions? What are the criteria for assessment and the principles for designing such assessment in these conditions?

1.5   Aims of the research

The aims of this research project were to:

(i) understand the performance features of the Clinical Skills Assessment (CSA) and, in particular, to investigate the extent to which linguistic/cultural factors contribute to poor performance, including the specific conditions of the exam which produce this performance. The project was catalysed by the observation that International Medical Graduates and Black and Minority Ethnic Group UK -trained graduates performed less well than white UK-trained graduates, as demonstrated in the MRCGP Annual Reports. This differential pass rate was also seen in the Applied Knowledge Test (a machine marked anonymised multiple choice paper), but the differences were magnified still further in the CSA.

(ii) develop an analytic framework and design associated materials to raise awareness among examiners, GP trainers and candidates of the linguistic and cultural demands of the exam.

This project (2011 – 2013) was part of the UK-wide Knowledge Transfer Partnership programme, funded by the Technology Strategy Board, with additional funds, in this case, from the Academy of Royal Medical Colleges. It was set up as a collaboration between the RCGP and King's College London with support from Cardiff University (see Acknowledgments). The rationale for this joint project was that a multi-disciplinary team of GPs and linguists working together could draw on each others' expertise to explore a widely recognised problem in new ways.

1.6   Background to the project

The partnership between King's and the RCGP goes back to the mid-1990s when the RCGP commissioned a small piece of research to explore the gap between the success rates of overseas trained doctors compared with those trained in the UK in the RCGP membership examination (Roberts, Sarangi, Wakeford, Wass and Southgate 2000). The KTP was set up to tackle a similar problem but with a more high stakes examination, the CSA, which licenses doctors to work as GPs in the UK. Since the exam was first run in 2007 there has been a gap in success rates between UK-trained and International Medical Graduates: See appendix A for a summary of the history of the development of the CSA. The latest pass rate figures are published in the MRCGP Annual Report (2013). In 2012, over 3000 candidates took the Applied Knowledge Test (written paper) and Clinical Skills Assessment (clinical OSCE-style examination). For the AKT, 87.7% of UK graduates (UKG) passed at the first attempt and 50.3% of International Medical Graduates (IMG). For the CSA, 90.1% of UKG passed at the first attempt, and 34.7% of IMGs.

Introduction

The College's long-standing commitment to tackling inequalities has led to the publication of CSA results categorised by place of training. This transparency has, in turn, led to increased concern about what action could be taken to close the gap and how to respond to the reactions from GP trainers, IMGs and overseas doctors' associations.

While several general reasons have been suggested for this persistent gap (Woolf et al 2011, McManus and Wakeford 2014), no detailed study of candidate performance had been carried out. The KTP, therefore, proposed to use linguistic and micro discourse analysis methods to identify the performance features of the exam, to contrast the features of successful and failing candidates and the basis on which they were assessed. This partnership between the RCGP and King's College London, with Cardiff University, together with the RCGP's publication of a detailed MRCGP Annual Report containing equality and diversity data, is an indication of the college's willingness to be transparent and its intentions to conduct a fair but rigorous clinical examination that adds to the assessment of a general practitioner who is fit for independent and safe family practice in the UK.

1.7   Brief history of the CSA

As described in Appendix A, the Clinical Skills Assessment is one of a tripos of assessments introduced in October 2007 after a two year period of design and preparation. It uses the principles of an objective structured clinical examination (OSCE) to test a number of competencies from the MRCGP curriculum that can be tested in a simulated surgery. These include patient-centred care, problem-solving skills, a comprehensive approach, community orientation and a holistic approach. Each of the thirteen cases is marked by a different examiner, but all cases have a generic marking schedule which comprises of three domains: data gathering (getting information from the patient's history and examination as appropriate and the case notes), clinical management (in line with usual NHS general practice, and using evidence based medicine where possible), and interpersonal skills (the doctor/ patient relationship where the doctor is expected to identify and respond to the patient's ideas, concerns and expectations). The CSA has been defined as 'an assessment of a doctor's ability to integrate and apply appropriate clinical, professional, communication and practical skills in general practice'. It is intended to be a means of assessing a doctor's ability to synthesise and assimilate information from the patient and the patient's case notes, and then to apply this to the problems presented in the 'case', taking into account the concerns and ideas of the patient, in a variety of clinical contexts taken from British general practice. The rationale for this is that a GP's work centres around surgery consultations, for which fluent, patient-centred consulting skills are needed.

Chapter Two.
Data and methods

2.1   Summary

This study is primarily a linguistic analysis of what candidates actually *say* in the CSA. This is important as the first evidence-based look at what happens in the exam itself using sociolinguistic and corpus linguistic methods, in an attempt to better understand why the disproportionate pass rates, shown by more general statistics, might be occurring. As outlined during this chapter, this linguistic aspect of the project comprises of:

- Collecting videos of consented candidates sitting the exam and carefully selecting a sample of 40 videos for close analysis.

- Conducting quantitative, 'corpus linguistic' methods (see Linguistic terms 1, next page) to get an overview of how candidates talk (**Chapter 3**), as well as a systematic, micro linguistic analysis of talk at localised levels (**Chapter 4**). Both approaches illuminate features of successful/unsuccessful interactions.

However, in order to get the best 'inside' view of how the CSA works, a range of ethnographic methods were used to support the core linguistic analysis. These included:

- Gaining as much contextual information on the 40 candidate 'cases studies' as possible, including the consented background information available on the candidates, as well as examiner checking of cases and their marks.

- A qualitative thematic analysis of examiner video feedback sessions on candidates' CSA videos (**Chapter 5**)

- An analysis of the marking schedules and particularly the criteria for the 'Interpersonal Skills' domain within the 40 cases in the sample (**Chapter 6**)

- Other ethnographic data to help inform an understanding of the CSA; reading and discussing marking material with examiners, being involved in the running of the exam, from work on the reception desk to meetings with the CSA core group. We also met with other researchers working on the subject, including Mohanna (2011) and Foreman (2013).

It is important to emphasise that these different layers of contextual data were used to inform decisions about further linguistic analysis and to corroborate the analysis in the later stages. Such use of ethnographic data is common practice in sociolinguistic studies and ensures that the analysis of interaction remains embedded in the processes and structures of the institution which has come to produce such interaction. In this instance, ethnographic data allowed the research to be more accountable to the overall exam process.

From this research, it was intended both to develop e-learning materials that provide candidates with a practical, evidence-based look into the communicative requirements of the exam, as well as analysing performance factors and the exam context from which they arose.

Details of cases cannot always be given for reasons of confidentiality.

Data and methods

2.2 Introduction

Rather than thinking about language as a set of forms (grammar, vocabulary etc.), language is social conduct. It is what we use to get things done and form social relationships on a daily basis. As such, analysing language can offer valuable insight on how the social world functions. Linguistic analysis is a broad-based approach which, for the purposes of this research, includes discourse analysis, micro-interactional analysis and more quantitative 'corpus linguistic' approaches (see Linguistic terms 1). These quantitative and qualitative methodologies all involve a systematic study of language data, but yield different perspectives. Quantitative techniques, such as corpus linguistics, used in Chapter 3, give us an overview of linguistic data and can answer questions about the general characteristics of talk in a particular setting. Qualitative techniques, such as micro-interactional analysis used in Chapter 4, look closely at data to unpick how an utterance functions in the particular context of use. However, there is no simple equation between macro and quantitative and micro and qualitative. In this research, quantitative and qualitative methods are used at both macro and micro levels.

While linguistic analysis provides a new lens for looking at interactions such as the CSA, the linguistic performances alone cannot illuminate the whole process of the exam. A range of ethnographic methods in this study opened up the focus to include the exam processes and examiners' judgements. As part of this, we held 'examiner feedback' sessions to discuss video clips of candidates. We also drew on the expertise of an advisory group of CSA examiners and GPs. We were firmly embedded in the processes of the MRCGP, over a period of two years, allowing us to better adjust to the perspectives of the institution. This included participating in the administrative processes of the exam itself, such as working on the candidate registration desk, as well as attending internal meetings and conferences for the MRCGP, talking to examiners, RPs, GP trainers candidates and researchers.

> ### Linguistic terms 1
> ### What's corpus linguistics?
>
> Corpus linguistics takes large datasets of 'real-world' language (called a 'corpus'), to make quantitative analyses of the patterns that occur. It's very useful in establishing the frequent words and phrases people use, as well as the immediately co-occurring language (a 'concordance').
>
> The idea of counting the frequency of words and their contexts has been around a long time (500 monks manually produced a concordance of the Bible in 1262). But it's become a widely used method with the advent of computers and the ability to make quick statistical calculations. It's good at testing our intuitions about how datasets of language in specific contexts such as healthcare (specialised corpora) differ from other contexts.

This research did not include interview or questionnaire methods. While these can be valuable in many research contexts, there remains the perennial problem of the gap between what people say and what they do (Brunsson and Jacobssen 2000). The time consuming and laborious process of recording what people actually do can be very insightful.

2.3   Data collection

Analysing what people actually do requires some empirical language data. The key data collection for the study involved:

- Videos of 200 informed and consented candidates sitting the CSA (in total just over 2500 cases), filmed during two diets in 2011 (28 February-11 March and 17-23 May 2011). After two cases were withdrawn (see ethics process below) the video data bank consisted of 198 cases.

From these we selected:

- A purposive sample of 40 cases to transcribe in detail as the key focus for linguistic analysis.

- Any additional cases tagged as 'Promoting equality and diversity' or 'Diversity'.

- Marks and demographic information for all 198 candidates.

- A case book of detailed feedback both from the exam mark sheets and additional examiner viewings for the project.

Additional contextual data:

- Transcripts and notes from 4 examiner feedback sessions – discussing clips of CSA candidates

- Marking schedules and RP/examiner/candidate notes for the 40 cases sample

- Statistical modelling data on the domain marks of a one cohort of candidates from February 2012, analysed by Richard Wakeford, the MRCGP CSA psychometrician.

- Analysis of the three domains used in case specific marking of all the cases represented in the 40 case sample.

- Ethnographic notes – from participating in the running of the CSA and meetings, conferences and conversations with examiners and others throughout the project.

- Other research and development work undertaken by RCGP members.

2.3.1   Ethics process

While 9 of the 39 CSA rooms were already filmed at the time of data collection, the video tapes are not routinely stored or used for research without specific informed consent from the candidates. The process for this was discussed with the CSA core group, the CSA administrative team and chief examiner, and ultimately approved by research ethics at King's College London. The chief consideration was to develop an informed consent procedure for candidates, allowing them to opt into the research project, while not causing undue stress prior to an important examination. For this reason, various stages of consent were offered. Similar consideration was given to requesting actors' permission. Through the various routes offered, 34 candidates

Data and methods

declined, but a further 198 (originally 200 before retrospective withdrawal) consented to have their video footage used in the research.

2.3.2   Data – selecting a 40 case sample

Central to the study was the selection of cases to be transcribed. This was to be a non-probability based, purposive sample. evidencing communicative styles relevant to performance in the exam. The sample was to encompass:

1. the range of candidates' training backgrounds, both UK and non-UK

2. the diversity of candidates according to their gender and declared ethnicity

3. high performing and low performing candidates

A sample of high scoring and low scoring candidates was selected based on overall pass mark as well as declared information on sex, ethnicity and place of medical qualification (see Appendix Table B-2). This involved a degree of essentialising into 3 groups – international medical graduates (IMG), UK ethnic minority graduates (BME) and UK white graduates. 'IMG' was considered to be any non-UK medical graduate, so our final selections of 17 IMGs includes 4 people who originally qualified in the EU. This grouping ties in with current statistical research on the exam (e.g. Denney et al 2013), where these divisions are necessary due to the actual numbers of candidates from different ethnic groups. Equal numbers of male and female candidates were selected. Given the disparity in pass marks it became problematic to find high and low scoring candidates with comparable marks across all groups.(Appendix Table B-1 illuminates this skewed range). In fact, we had no UK white candidates who failed the exam in our sample of 200. We therefore had to take the lowest scoring white UK candidates. This may be a weakness in our approach, since 'failing' cases may not mean they have failed the exam overall. So the performance gap between successful and unsuccessful candidates would have been somewhat narrowed. Nevertheless, this is the closest we could get to representations of poorer performance in this group.

Cases were selected for review which were roughly representative of those candidates' performances – i.e. 'failing' cases for lower scoring candidates, 'passing' for higher scoring. An equal number of 'failing' and 'passing' cases ticked with the 'borderline' standard setting were chosen so that we had a range of performances, not simply outright passes or fails. Through this process it was hoped all groups of candidates could be observed in poorer cases. We also selected 2 additional IMG cases from 'marginal failing candidates' (2-3 marks from passing the CSA). Cases went through a review before appearing in the 40 sample. Failing cases were sent, with the marking schedule, to CSA examiners on our advisory panel to ensure they had been reliably marked (see Appendix Table B-3).

As far as possible, cases were excluded where there were clear clinical errors or omissions. The marking schedule does not distinguish between poor data gathering and clinical management because of poor communication/IPS or because of errors/omissions. We took guidance,

therefore, from the panel to exclude those likely to have been marked down for medical reasons. Since this is a qualitative and micro-analytic study, the goal was not to achieve absolute representativeness but to use a reasonable cross section of candidates and to investigate the detail of their performance in order to understand it and how it is assessed. Cases were selected from the full range of 198 cases video recorded for the research and represented a mix of routine and complex cases, both cases with a predominately medical focus, and cases with a predominately psycho-social focus. Complex psycho-social and emotionally demanding cases made up just under a fifth of the dataset (this maps to the case selection palette for choice of cases for each exam circuit which states that 2-3/ 13 cases on the circuit should have a psycho-social focus).

A detailed 'case book' of the 40 cases was made during the research which specified the particular marks for the CSA, the feedback statements they received for the case, their AKT result, how many times they had taken the CSA and demographic information. Each of these 'case studies' also included the marking schedule, role-player's (RP) notes, candidate's notes and examiner's notes, as well as our own analytic memos and examiner feedback on the case. This reference guide allowed us to amalgamate the different perspectives gained from our various data collection techniques and we used this to inform and guide us through the detailed linguistic analysis and comparisons between candidates.

2.3.3   Data – examiner video feedback sessions

Alongside the data collection of videos, we conducted 4 sessions with CSA examiners during 2011. These showed clips from a range of CSA cases (both from the 40 sample and from the full 198 videoed cases) and examiners had the opportunity to discuss how they reached conclusions about a candidate's performance. Some direction was given in the session notes to; '...focus on what it is about the candidate's performance that you evaluate as good or poor and what particular aspects of their consultation lead you to this conclusion'. As a research team, these discussions allowed us to clarify more opaque (but often used) comments about a candidate's communicative performance, such as 'clunky', 'formulaic' or 'cold' with greater detail. These feedback sessions were transcribed and a thematic analysis was undertaken. The results are outlined in Chapter 5 and informed directions for our linguistic analysis. As with any group discussion used for research purposes, such as examiner video feedback, there are limitations which may underestimate the degree of heterogeneity and these are considered in Chapter 5.

2.3.4   Data – the marking schedules

In order that cases are marked consistently, there exists, for each case, a detailed set of descriptors of good and poor performance. We conducted an analysis of case descriptors in their own right to better understand the way domains in the CSA (Data-gathering, Clinical Management and Interpersonal Skills) were assessed. This included an appraisal of 'Interpersonal Skills' (IPS). The first step mapped the presence of IPS statements in the

Data and methods

following CSA paperwork:

- Case specific marking statements, as divided into the three marking domains of Data Gathering (DG), Clinical Management (CM) and Interpersonal Skills (IPS).

- The actual distribution of case specific marking statements within marking schedules for the 40 cases that are analysed for passing and failing candidates.

These results are discussed in Chapter 6, alongside the examiner feedback sessions.

2.3.5   Data – diversity cases

The blueprint for cases for the CSA circuits includes the criterion that at least one per exam must include a significant 'diversity' element based on the protected characteristics under the Equality Act 2010. In addition to the transcribed 40 case sample, we therefore also took any that represented 'diversity' from the 198 cases. These were used to analyse the types of communication in diverse settings they represented. Cases either tagged as 'Promoting Equality & Valuing Diversity' in the CSA case bank or those that had been assessed as 'Diversity: Ethnicity' and 'Diversity: Culture' in a previous analysis of the case bank by the Case Management Group were identified in all those recorded during our CSA data-collection. From the 198 separate cases we had on film, 4 were tagged as 'Promoting Equality and Diversity' and a further 5 indicated as 'Diversity: Ethnicity' or 'Diversity: Culture' cases (see Appendix Table B-4).

2.4   Transcribing the consultations

To make a linguistic analysis of 40 cases, it was necessary to transcribe them. Transcription of spoken interaction is never a straightforward representation of what is 'said' but always involves making choices about what will be represented; features such as intonation, speed of speech delivery and gesture can be represented as well as words. For this project, since examiner feedback sessions frequently commented on the non-verbal, we wanted to continue to analyse these as videos, rather than just textual transcripts.

We made the decision to use a 'video-linked' transcription software, CLAN ('Computerized Language Analysis' 2013) enabling quantitative analysis and the close investigation of the transcript text alongside the linked video. After converting the original CCTV footage, the transcription process then involved creating an audio file, and linking sequential sections of the waveform to transcribed dialogue (see Figure 1).

Figure 1 – Transcribing the consultations
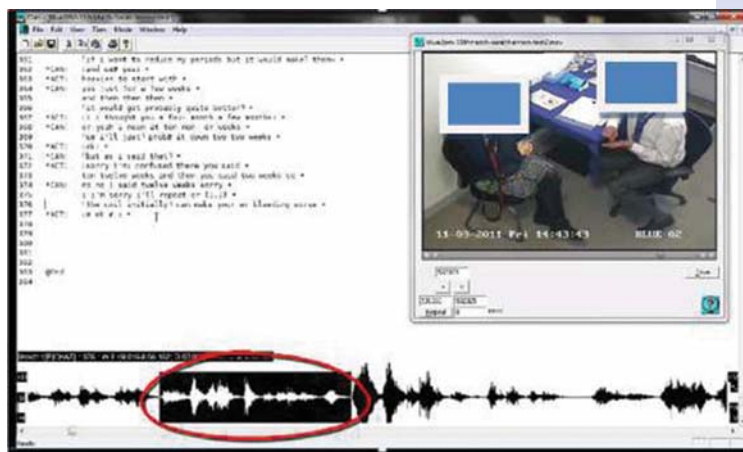
2.5 Mapping and annotating the transcripts

Once transcribed, our initial analytic process was to annotate the transcripts for structure and key communicative features, producing a visual 'map' of each consultation. These annotations went through several iterations as CR and SA discussed and tested them for consistency on increasing numbers of cases. The eventual annotations decided on represented the most recurrent and important features. This was done initially by looking at the transcripts 'blind' – i.e. without information on a candidate's marks or background. This helped mitigate against the confirmation bias that can result from analysing cases known to be poor or high performing. We mapped the broad phases of cases:

- **Data-gathering** – dialogue eliciting information from the patient.

- **Physical Examination** – even if it was simulated (i.e. a card was handed over by the examiner).

- **Explanation** – where a decision is conveyed by the candidate to the patient or where a particular medical issue/treatment is elaborated on.

- **Discussion of action** – all dialogue involving reaching a decision about what to do next – whether this be treatment, referral, ethical decision etc.

- **Key-problem discussion** – a feature of more 'complex' cases, where the crux of the cases rested on a key ethical or emotional issue – a key problem discussion identified and talked about with the patient, without necessarily taking a particular position or decision.

These phases could last for extended periods of the consultation. In addition to this, we also annotated more micro, local level moments of interaction that were important to the consultations, but took place only over the course of a few speaker turns. These were:

- **Role-player sink/save** – moments where the RP appears to help or hinder the candidate in the success of the case or the success in creating mutual alignment.

- **Prefacing** – where a candidate introduces an idea gradually, with an earlier hint towards this in their dialogue, as well as overt sign-posting of things to come in the consultation

- **Alignment** – expressions of overt agreement, compliance or moments in the consultation where the dialogue between candidate and RP is flowing well.

- **Misalignment/Misunderstanding** – moments of overt misunderstanding, communicative 'trouble' between candidate and RP, expressions of disagreement.

- **Exam Modelling** – use of the phrases identified as CSA 'formulaic' utterances or CSA 'exam-techniques' – i.e. those features identified either in examiner feedback sessions or in the corpus linguistic analysis (3.4) as being indicative of a CSA exam model.

Data and methods

Case Types – As part of the mapping, we categorised cases as to whether they presented complex or routine challenges (see Appendix Table B-5). Complex refers not to the difficulty of the clinical problem but the complexity of the interaction, apparent from mapping their structure. These usually involved a prolonged period of agonistic dialogue (see Linguistic terms 2) with the patient, which teased out the institutional, professional and personal tensions with the patient's request.

Linguistic terms 2 Agonistic dialogue

'Agonism' refers to the positive aspects of intractable political conflict but it's also a useful concept for thinking, at this more micro level, about how power and institutional tensions play out in talk. '**Agonistic dialogue**' is used here to mean interactions where speakers are negotiating incompatibly different values, from which a consensus decision is not entirely possible. For example: the mother in the *'MC 1'* case feels it's right to give her son medication without his knowledge, a view the doctor cannot align with. To gain a good mark, successful candidates must tease out these discordant values (agonistic dialogue), balancing a professional stance, on the one hand, against alignment with the patient, on the other, without breaking down into an out-and-out fight (antagonism). This is a complex 'rule of the game' that goes beyond any simple notion of patient centredness. In many linguistic and communication models, successful doctor-patient consultations are usefully thought about in terms of a goal-oriented interaction. But in these agonistic cases, that notion of a 'common goal' becomes fuzzier.

Such cases, as in the *'MC'* case, usually present an initial request:

```
RPL: i've come about my son john
CAN: alright
RPL: i'd be very grateful if you could er (0.2) change
his medication from tablets (0.7) into ones that
dissolve
```

but quickly follow this with a hint towards the crux of the case:

```
RPL: so (.) i decided the only thing to do was for me
to (1.7) well (0.5) for me to intercede and i would
(0.7) breaking up these tablets which is quite hard to
do hhhh erm (0.3) and sprinkling them on his (.)
breakfast (.) cereal
```
(*'MC 1'* Case, UKG)

Rather than provide an explanation of the medication here, these types of cases involve the candidate correctly inferencing the moral problem at stake. As discussed in examiner feedback sessions, cases often do not present a straightforward resolution or even require a final decision within the 10 minutes. What they require is the identification and sensitive discussion of the problem in a way that does not break down in conflict. There is not necessarily a 'right' answer, but the candidate is expected to show a mode of approach that is justifiable. As such, they often presented different linguistic features.

2.6   Overview of the linguistic analysis

The 40 transcribed and annotated cases were then analysed in the following ways, which comprise the structure of the following 2 chapters:

Chapter 3: 'Broad based' analysis:

- Timings – structuring cases (how long spent on data gathering, explanations and discussion of treatment).

- Method – Using the annotated transcripts, already time stamped in the CLAN software, to automatically calculate the length of phases and represent the global structure of consultations.

- Timings – who speaks when and 'floor holding'.

- Method – Using the time stamped transcripts in CLAN to calculate the length of speaker turns, the number of interruptions and overlapping talk between speakers, and how this changes through the different phases of the consultation.

- Corpus linguistics – the words and phrases of the CSA.

- Method – counting words and phrases speakers use through computer aided techniques and statistically comparing them to large reference banks of English in use.

Chapter 4: Micro-analysis

- Interactional and affiliative alignment – how speakers bring the talk into line with one another to achieve a common goal in the consultation.

- Formulaic talk – what aspects of delivery might cause some formulaic talk to seem appropriate in the CSA and some not.

- Explanation phases – how are they structured at the micro level, what linguistic strategies candidates use to convey medical information to the RP. How do they create a sense of dialogue in this phase in which it is largely the doctor who talks? What other features of talk eg intonation accompany the delivery of medical information?

The combination of these approaches enabled us to triangulate findings from different aspects of research to inform our conclusions. The linguistic features, particularly the patterns of similarity and difference between unsuccessful and successful candidates, will be picked up in later chapters in the wider discussion of implications.

Data and methods

2.7　Mixed method and ethnographic analysis

This research, as outlined above, used a mixed method approach for understanding interaction and oral performance. Drawing on both quantitative and qualitative sociolinguistic approaches at two levels: a relatively broad-based level looking at patterns across the whole data set and an in-depth, detailed level (micro level). These core data were supported by ethnographic data to give more analytic precision and connect the interactional analysis to the wider structures and processes of the exam. There is no simple equation between macro and quantitative and micro and qualitative. In this research, quantitative and qualitative methods were used at both levels. These approaches and levels are complementary and necessary to achieve the goals of the project.

The hallmark and explanatory power of much sociolinguistic analysis lies in its tools for exposing and understanding the teeming life which lies within talk. Micro-analysis, therefore, which forms the bulk of the analytic work, provided some of our most useful findings. However more broad-based analysis was important in identifying interactional patterns which needed to be researched in more depth and in establishing the context for our analysis at the micro level. This broad-based analysis, resulting from coding and quantitative analysis of transcripts, identified certain performance features of candidates and examiner reactions. This analysis, while telling about the nature of the exam, is somewhat limited because of the relatively small database.

The unit of analysis at the micro-level is much smaller (i.e. at the level of word, phrase and sequence of interactive turns) and so provides a larger data base at this level for both qualitative and quantitative research. It is used to collect and unpack data in very different ways from traditional methodologies and so reaches levels of analysis that these other methodologies cannot reach.

Firstly, this small unit of analysis means that relatively large numbers can be used for computational micro-analysis, as described above. Secondly, qualitative micro-analysis uses an aesthetic of smallness and slowness to:

1. understand how talk, moment-by-moment, is produced and determined by a particular context and, especially in high stakes assessment, how small differences or discrepancies themselves produce a new context by which speakers may be judged.

2. identify the conditions for talk, even in a single telling moment, such as a misunderstanding, which are generalisable across broadly similar settings (such as OSCE type settings). For example, our analysis explains why, although all candidates use 'empathic' phrases, poorer performing candidates are seen as formulaic and so less interpersonally effective because the conditions of such an examination require a demonstration of 'empathy' spoken in a particular style.

The combination of different types of linguistic analysis enriched by ethnographic data (i) helps to explain and offer solutions to interactionally based social problems (ii) sets detailed analysis

of data within the specific conditions of their production so that previously hidden relationships and phenomena become apparent (iii) allows for applicability of concepts, patterns and phenomena to other contexts, provided these contexts meet the same conditions as those in the original data (iv) stimulates creative thinking and can disturb dominant assumptions.

In sum, the methods used provide 'telling' cases (Mitchell 1984:239) which identify patterns of interaction and their relationships to the wider social conditions of the exam. What gives the findings the potential for more generalisable comment and application to other contexts, therefore, is not how far they can be said to be statistically representative of all candidates, but the extent to which they build new says of seeing and thinking which are relevant to similar contexts.

There are, however, challenges that can be made to this research. Most notably, these stem from fundamentally different traditions of how knowledge comes to be produced through research. Positivist traditions assume that useful knowledge is the product of statistically quantifiable evidence. Interpretive research, such as this project, is valued on how far it increases understanding. A mixed method approach which uses some quantitative aspects but is largely qualitative and interpretive is vulnerable to criticism from both traditions. Our goal was to be as holistic as possible and avoid decontextualised abstractions, on the one hand, and over-specificity on the other, but within the time constraints of the project, some limits to this mixed method approach are inevitable. The database of 40 cases is relatively small for undertaking corpus linguistics and for the coding of different phases of the simulated cases. By contrast 40 cases is a large data base for the very detailed micro-analysis we undertook and so we opted for looking at certain features which previous intercultural research had identified as particularly salient. The data set from the ethnographic feedback with examiners was also small and the analysis of such elicited data was subject to the usual constraints of such data (see Chapter 5). However, this data set was to support the core linguistic data rather than being the major focus of the project.

Chapter 3.
Broad analysis of the 40 cases

3.1   Summary

This chapter outlines findings from quantitative investigations of a sample of 40 cases. It looks at the structure and timing of cases, the talking time of speakers, the most frequent words and phrases used and frequency of interactional features. At this surface level, the close similarities between candidates, whether passing or failing, is striking but some features do emerge with implications for the e-learning materials and the exam:

Structuring and timing

- In routine cases, there is little difference between successful and unsuccessful candidates. It seems many candidates understand well the typical model for structuring consultations.

- Late data gathering, after 'half time', is the only feature of poor performance and always contributes to a lower score. This usually occurs to repair missed information and can often still lead to the expected management plan.

- Complex cases are structured flexibly, with shorter data gathering and more time for ethical discussion. This highlights the need not to be too prescriptive in the e-learning material.

Talking time and interruptions

- Successful candidates in routine consultations talk for around 68% of the floor time on average. This is a little more than in 'everyday' GP consultations and to be expected in a spoken exam.

- Explanations are interactionally demanding, with candidates having to hold the floor for a much larger proportion of the time than in the consultation as whole.

- Greater flexibility is required in complex cases, particularly with judging appropriate opportunities to speak less and allow the role-player more of the floor time.

- There is a good deal of overlapping talk and interruption in the CSA, just as in any spoken context, but no difference between passing or failing candidates.

- Interruptions in the explanation phase are, though, problematic. They function to highlight inadequate explanations and disrupt the speaker's flow and speech planning.

Words and phrases in the CSA – corpus analysis

- There are some words and phrases that are very particular to the CSA when compared with spoken English and 'everyday' GP consultations, suggesting a CSA 'linguistic fingerprint'.

- Formulaic expressions cluster around the social/interpersonal work of the consultation – phrases using 'feel', response tokens 'OK'/'alright' and questions using 'any'/'anything'.

- There are some indications that CSA consultations reflect a more local English – they are marginally more like John Skelton's GP consultation dataset than another dataset of consultations in Lambeth where GPs were frequently interacting with patients with less than fluent English. Nevertheless the small size of the CSA dataset needs to be taken into account.

- There are very few lexical differences (differences in words/vocabulary) between candidates – but some differences that point towards features of local English, such as vague language hedges 'sort of/kind of' used more frequently by successful candidates.

Interactional features

- Successful candidates did show marginally more alignments and fewer misalignments. IMG candidates had slightly more misunderstandings.

- There was no clear evidence that role-player behaviour systematically led to success or failure.

- Successful candidates tend to use more exam modelled strategies. However these were the strategies deemed 'formulaic' in examiner feedback, a paradox further investigated below.

Overall, a lack of difference between candidates at this general level requires us to look more closely at the dialogue. General skills such as how to structure the consultation appear to be mastered well by the majority of candidates, and they all use similar professional language. So we must put the microscope on the interaction in Chapter 4 to understand why some candidates are much lower rated than the majority.

3.2   Timing and structure

With a strictly timed 10 minutes, time management is inevitably a driving factor for communication in the CSA. The 'mapping' of the 40 cases allows us to answer some key questions about time management raised by examiners in feedback sessions. There was a feeling that weaker candidates tend to spend too much time on data gathering: 'I think watching other candidates quite often they delay making a decision for as long as possible. And they spend ages on the data gathering' (Examiner Feedback Session 1). Nevertheless, the converse was also true that successful candidates were often seen to take a good amount of time 'working out what it is all about'. Spending the correct amount of time on data gathering, and how this appeared, therefore seemed to be a key issue to explore from the mapping.

3.2.1   Average time spent on phases

Adding up times spent on broad phases ('Data Gathering', 'Physical Examination', 'Explanations', 'Discussion of Action', and in some complex cases 'Key Problem Discussion') allows us to get a rough picture of how the consultation is divided up. Appendix Table B-6 in

Broad analysis of the 40 cases

the Appendices show the timings for low scoring candidates, and Table B-7 the high scoring candidates.

In routine cases, on average poorer candidates spent 3:29 mins on data-gathering, almost identical to the 3:33 average in successful cases. A similar picture emerges for time spent on 'Discussion of Action', taking up just over a third of the time in routine cases, leaving the rest for medical explanations, greetings and checks. It is notable that one of the poorer performing candidates, in *'SH 1'*, only spends 23 seconds on the 'Explanation'. Successful candidates, in routine cases spend around a minute or more on this. Nevertheless, spending too long on 'Explanations' seems to be the downfall for a UKG performing *'JH'*, who in fact has to backtrack and repair a mistake, leaving little time for 'Discussion of Action'. There is an occasional slow candidate in the poorer cases; notably the UKG performing the complex *'MC 1'* case, who spends 5min 44 sec on Data Gathering. However, the majority of low scoring candidates get through 'Data Gathering' in a time very close to high scorers. Overall, despite a few stand out disasters, what is notable is how similar time structures are, suggesting that many candidates, weak or strong, know how to manage structuring. Examiners' comments that weaker candidates spend too long data gathering aren't borne out here, but another pattern occurs, outlined 3.2.3 (late data gathering) that can account for the sense of poor structuring.

In 'complex' cases, less time is spent data-gathering to make room for phases comprising a back-and-forth discussion of the core ethical problem. While hard to generalise, this begins to give a picture of how these simulated cases are structured and shows that candidates need to be prepared to structure the case differently when presented with complex scenarios.

3.2.2   Switching at 'half time' in routine cases

The time stamped transcripts allow us make a graphical representation of where phases occur. This *'RS'* case (Figure 2 below) on a patient with a psychosexual problem of premature ejaculation is with a female IMG candidate who gained a 'Clear Pass' and an overall mark of 85 in the CSA. This structure can be seen as a typical exemplar for the majority of routine cases. The tendency is for all the data gathering to take place before the midpoint of the consultation (i.e. up to or at the 5:00 min marker), and then a mixture of medical explanations and discussion of action taking place in the 'second half'. The occasional short explanation or hint
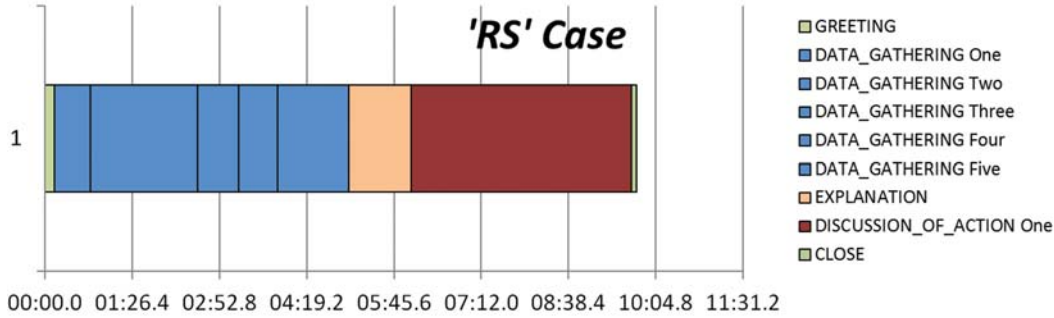
Figure 2 – Structuring of phases in *'RS'* Case

towards the treatment discussion can be offered early on, but any detailed discussion is held off until after 'half-time'.

3.2.3 Late data gathering

While the overall time spent on data gathering did not show much difference between poor and strong candidates in Section 3.2.1, the location of data gathering phases did play a role. Those candidates who carried out additional short phases of data gathering later in the consultatin, i.e. well after the 'half time' marker, such as the *'JL'* candidate (Figure 3 below), nearly all did badly, perhaps indicating something missed in the consultation or a disorganised structure.
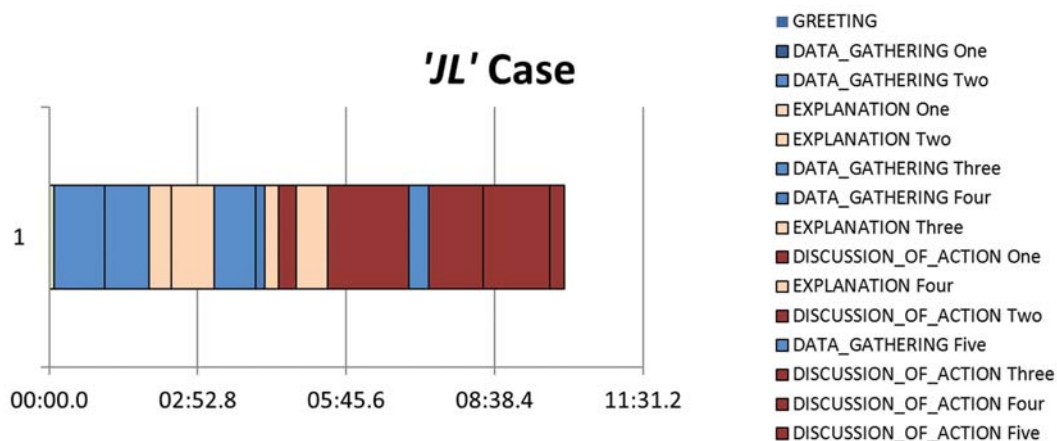
Figure 3 – Structuring of phases in *'JL'* case – late data gathering

The scores of failing candidates are in the range 56 – 71. The pass mark varies from within each diet of the exam and across diets but it is usually in the early 70s.

The IMG candidate scores 4/9 on the *'JL'* case and 56 overall in the CSA, with the feedback comment that he has a 'Disorganised/unstructured consultation'. Perhaps one problem here is that he is giving the RP results from a test for hypothyroidism, meaning that he gets into a medical explanation too early on, before fully establishing the RP's understanding of what the tests were for. Examiner feedback commented: '…the consultation was clunky. He covered some ground in the history more than once, and did not explain what ***** disease is at the appropriate stage in the consultation'. So, unfinished data gathering in the early stages has a knock-on effect on the staging of both data gathering and explanation.

Similarly late data-gathering phases can be found in many more of the poorer cases:

- *'JK'* – 3/9 Clear Fail, 57 overall

- *'MC 1'* – 4/9 Clear Fail, 71 overall and feedback 'Shows poor time management'

- *'MS 1'* – 4/9 'Fail' Borderline standard setting, 56 overall and feedback 'Shows poor time management'

Broad analysis of the 40 cases

- '*KM 2*' – 4/9 'Fail' Borderline standard setting, 68 overall, feedback 'Disorganised/unstructured consultation'

- '*MBS*' – 4/9 'Fail' Borderline standard setting, 57 overall, feedback 'Disorganised/unstructured consultation'

- '*ES 1*' – 4/9 'Fail' Borderline standard setting, 68 overall

This structural issue, and particularly the location of data-gathering phases, accounts for over a third of the poorer performing cases (7 out of 18), a much clearer pattern than any indication that poorer candidates spend too long on data-gathering to avoid decisions. If anything it seems important to spend the majority of time in the first half of the consultation data gathering to avoid missing something that will emerge later and requires further questioning. It seems important for some of the e-learning material therefore to focus on how to think through the structuring of phases and in particular how the data gathering will be systematically performed in the first half.

Mistakes are of course made by strong candidates too and even those who did relatively well would sometimes conduct a data gathering phase later in the consultation having forgotten something or picked up on a new piece of information. But even for these successful candidates, the mistake had an effect on their marks. No 'Clear Passing' cases in our sample carried out data gathering after 'half time' – all such cases were 'borderline': *'EB'* (5/9 Borderline SS, 86 overall), *'KB'* (6/9 Borderline SS, 87 overall) *'RT'* (5/9 Borderline SS, 91 overall). In *'RT'*, for example, the candidate forgot to ask a question and comes back to this at 6:43 mins, leading him to carry out a second physical examination and amend his diagnosis. While he still effectively 'passes' this case and does well in the CSA, it would seem to have some effect on his marks and the ticking of the 'borderline' standard setting, as well as his getting the feedback comment of a 'Disorganised/unstructured consultation'. While it is possible to recover from a mistake of this kind and still achieve the correct outcome, it seems difficult to recover marks. While a disorganised consultation perhaps cannot be highly credited, mistakes can occur and it is worth considering greater opportunities to recover these in the CSA. Being able to recover from a mistake is a skill needed for General Practice which perhaps candidates need to know how to do efficiently, as well as examiners be able to note as a skill. Further analysis on the repair of mistakes is addressed in 4.4.

3.2.4   Complex case structure

Complex cases, i.e. those requiring the discussion of a more agonistic (see Linguistic terms 2,) moral dilemma or ethical decision, could take on rather a different structure, to leave space for a dialogue around the key problem of the case. Nevertheless, a good third of the case still needed to be spent on deciding some kind of action, for example with '*TJ*' (Figure 4), a case on a man's wish for a secret vasectomy (9/9 'Clear Pass', 109 overall):

While often no absolute decision is made in complex cases, as in '*TJ*', the candidate still manages a long period spent on explanations and discussing possible action. Complex cases where the
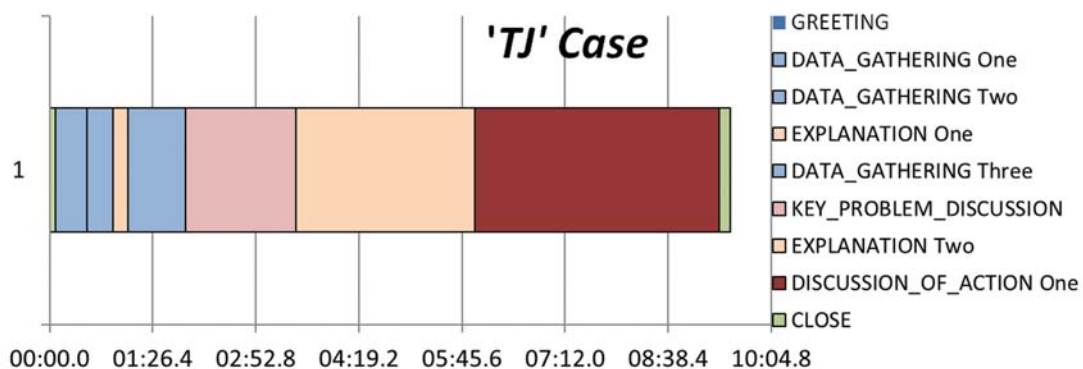
Figure 4 – Structuring of phases in *'TJ'* case – Complex scenarios

candidate did not get into this substantial discussion of action at all, such as '*MC 1*', were seen as too slow and indecisive, leading the candidate to do badly. A difficult balance has to be struck in these more complex cases, then, between spending enough time on the more discursive phases, but leaving enough time to bring the discussion back round to possible action, even if not completely resolved upon. Closer analysis in 4.3.1.1 also reveals how early decisions on the structuring of these complex cases must take place.

3.2.5 Overview of time/structure analysis – similarities between candidates?

Some interesting patterns emerge that might account for poorer performance in a few candidates, particularly in terms of the late data gathering stage. This has implications for training prior to the CSA. It is also interesting to note how differently these consultations become structured during more complex cases, with a greater need for flexibility. This again has implications for training, particularly in making inferences early on how a case is likely to pan out. But, overall, it is striking how similar the general structure and timings look for the majority of candidates:

The disrupted structure occurs in 7 out of the 18 poorer cases. In the other poorer cases, it is not possible to tell from the overall structure, what has caused a candidate to do well or badly. In Figure 5, '*SLY*', '*AH 2*' and '*SH 2*' are low scoring cases and '*RS*' high scoring in Figure 2, but it would be hard to tell from their structure. It becomes necessary, then, to look in more detail at what is actually being said in the consultation. In particular, our study identified issues with the way explanation *phases* were structured (see 4.6 below), as well as linguistic features and 'micro-moments' throughout the case.

Broad analysis of the 40 cases
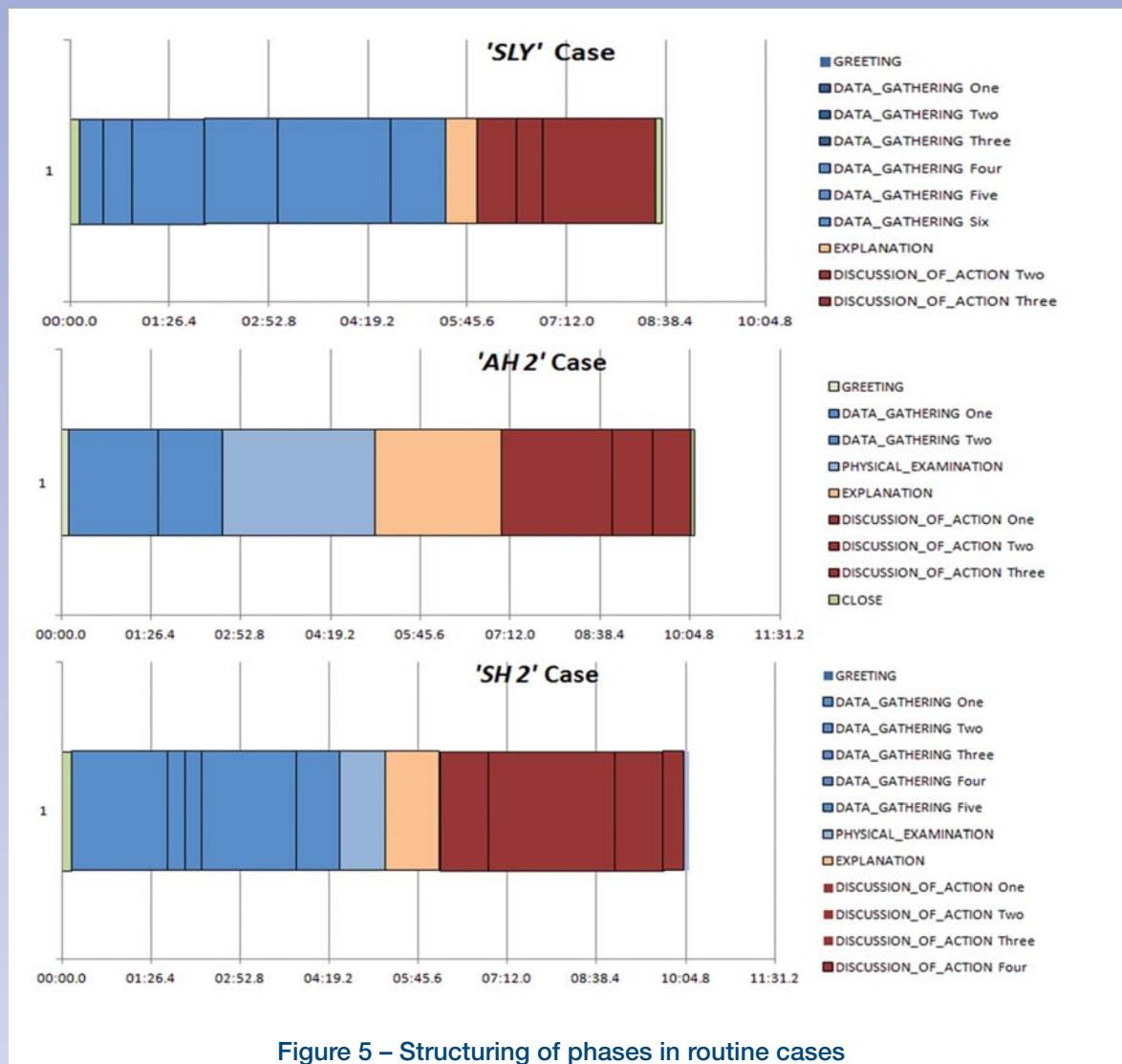
Figure 5 – Structuring of phases in routine cases

3.3   Broad analysis – turn-taking and holding the floor

Holding the 'floor', as a colloquial term, relates to who is speaking at any one time. Generally, one speaker holds the floor at a time, but interruptions and overlaps occur throughout talk as transitions between speakers. Although interruptions or holding the floor do not necessarily indicate interactional dominance, a pattern emerges in medical consultations which suggests, the doctor is expected to hold the floor slightly more than the patient. Roter (1989) suggests the doctor's talk makes up 60% of the consultation and Sanden et al (2001) that 'the ratio between doctor and patient as regards amount of talk is fairly stable'. Floor holding, interruptions and overlaps, (who talks where) can be analysed through the time stamped transcripts.

3.3.1   Analysis of speaking times

When all the time stamped consultations are looked at, it becomes evident that generally

candidates talk for the majority of the time (Appendix Table B-8). There are differences between ROUTINE and COMPLEX cases. In the successful ROUTINE cases, candidates range between holding 60-72% of the floor time when they're talking across these consultations, averaging out at about 68% of the consultation. It's worth noting that this is often, therefore, slightly more than the 50-60% identified in real-life medical consultations, which is perhaps to be expected in an exam that focuses on testing specific competencies through role-play, in which actions must be articulated to the overhearing examiner. In the ROUTINE cases where candidates do not perform well, while there is not a great deal of difference in the average times candidates talk, the variation between these is much greater, with candidates spending down to 54% and up to 83% of the consultation talking. While the amount of time spent talking in itself is not a hard and fast indicator of success, it does seem to be desirable to stick approximately to the 68% proportions. CSA consultations while desirably patient centred then, must still be largely dominated by the GP's talk.

What is clear again is the greater flexibility needed in the more complex cases, where even the successful candidate's talk ranges between 45% and 76% of the consultation. It seems that in these more difficult cases, the candidate must sometimes talk less. Nevertheless, this is perhaps still within certain parameters as the UK Graduate who fails the *'MC 1'* case and is seen as being too slow from examiner feedback, only holds the floor for 36% of the consultation, with the RP dominating. So while it may be acceptable to take more time and hold less of the floor in complex cases, there appears in these data at least to be a tipping point that will jeopardise the result.

3.3.2   Analysis of Interruptions

It also becomes clear that there is a high amount of overlapping talk (5th column in Appendix Table B-9) i.e. where the RP and candidate are talking at the same time. As a quick way of analysing this, using the timings, we automatically calculated the two following types of overlapping/interrupting talk from the candidate and the RP,

1. a simple interruption **–** A talks in B's turn, B stops, A takes over turn – simply labelled 'interruption' in our data. These are potentially dominating interruptions, where the interrupter gains the floor.

2. an internal overlap – A talks in B's turn, B keeps talking, A stops – we call these 'backchannel' overlaps, and they are often seen as a more supportive, non-floor holding type of overlap. (Adapted from Ferguson 1977)

Overall the numbers, again, do not show a great deal of difference between the passing and failing cases – all cases experience overlapping talk but there is no clear picture that either role-player initiated or candidate initiated interruptions make a difference to the outcome. What is significant is the relatively high level of role-player interruptions. This is quite different from research on 'real-life' consultations, where it has been suggested 67% of interruptions in a consultation were initiated by doctors, only 33% by patients (West 1998: 346). No such regular difference occurs here and if anything the two speakers echo each other – i.e. if one speaker

interrupts frequently, it is highly likely the other speaker will be interrupting a good deal. This amount of RP interruptions compared with 'real' consultations is also an indicator of the relative power of the role-players.

Interruptions during explanation phases do seem detrimental. High scoring candidates speak most of the time during explanations – between 72-95%, much more than the consultation as a whole, with fairly minimal responses from the RP. This is a particularly difficult phase for the candidate then, since they must do the brunt of the talking work, performing the difficult task of tailoring medical information to the particular needs of the patient and avoiding losing the floor. This is no small task, since it is difficult to 'plan what to say and speak at the same time' without sometimes stopping and replanning (Stenström 1994). Stopping or stating things in an unclear way is likely to lead to interruptions. Explanations were identified for closer analysis (Section 4.6), to develop training materials that assist the candidate to navigate and structure this difficult phase.

3.4 Corpus linguistics – the words and phrases CSA candidates say

This section uses corpus linguistics to overview the 'words' candidates use. This is a quantitative technique that can identify which words occur statistically more frequently in certain settings than in others, which words 'stick together' and who says them. There was often a sense in examiner feedback that candidates, particularly poorer ones, learned a grating set of stock 'CSA phrases':

> …because I thought that he was using *some formulaic words* but not in a formulaic way. So he does say, 'What did you have in mind' which is *one of those formulaic things people say* and the other one was '*How does that make you feel*' which is *another one that we pick up on a lot. But the way he said it was actually quite acceptable*. It didn't bring out antibodies…

A corpus linguistic analysis helps to shed some light on what people say and 'formulaic phrases'.

The 40 transcribed cases gave us a dataset, or 'corpus', of just over 85,000 'words' (this includes every utterance made by candidate and actor, down to every false start, 'um' and 'mmm' that occur). While this is relatively small for a linguistic corpus, in a very specific context such as this, around 80-100,000 words is enough to begin to find patterns. Furthermore, small specialist datasets such as this can be compared to much larger 'reference corpora', usually millions of words, in order to identify patterns and differences that we can say, with some statistical confidence, are characteristic of the specialist setting being examined. We compared the CSA dataset with some much larger reference corpora of general spoken English and also 'real-life' GP consultations (3.4.3), using a 'keyword' analysis. (For explanation of the calculation, see Appendix Definition – B-10) to highlight the words and phrases that occur significantly more or less frequently in the CSA than one would expect by chance.

Linguistic terms 3 What's 'formulaic language'?

In linguistics, 'formulaic language' means words that 'stick together', which we store in our memory whole, almost as though they're one big word (Wray 2002:9). These range from archaic idioms ('kick the bucket' (illustrated here)) to frequent, 'preferred ways' of saying things ('…you know what I mean'). Much of our everyday language is made up of these pre-composed, sequences, with calculations as high as 52-58% (Erman and Warren 2000). It's seen to be a positive feature of native speaker proficiency but we're not usually conscious of these stock phrases while we're speaking and listening. In the CSA though, some of the frequently used phrasing becomes consciously registered as sounding 'learned', 'formulaic' and grating – and therefore a negative rather than positive feature of talk. Expressing 'empathy' does seem to make pre-composed language sequences tricky: 'Repetition of a learned and rehearsed response or confining 'empathy' to a specific phase can communicate the exact opposite of empathy' (Roberts et al., 2003). Given how common they are in everyday speech, it's not surprising all candidates use formulaic sequences. The difference is which candidates sound '*marked*' in doing so.

3.4.1   'Keywords' – differences from general spoken English

We used the British National Corpus, spoken component (1 million words of 'everyday' spoken English) to compare with the CSA. Many of the 'single', stand alone keywords that surface (see results Appendix Table B-11) are incidental to the particular medical topics (e.g. 'gene'). Nevertheless, we can also see frequent words in the CSA that seem likely to be performing a more interpersonal function. In particular 'feel' appears in the top 10 keywords for the CSA, as do 'OK' and 'alright'. When we look at the words that frequently cluster together (Appendix Table B-12), then nearly all the topic specific keywords disappear and the remainder seem oriented to this 'interpersonal function'. Particularly noticeable are all the response tokens ('Ok Right Ok, 'Right Ok and 'Right Ok Ok'). In fact the frequent use of 'ok' was commented on in examiner feedback as indicative of CSA talk, a feature that does seem borne out in the data. Other clusters similarly point to 'CSA phrases' that examiners noted in feedback sessions; 'do you feel' as part of 'how do you feel about...' type questions; 'any/anything' questions; asking permission questions such as 'can I ask you a few...' usually 'personal questions'; phrases around 'how can I help you today' and 'can you tell me a bit more about' as openers were also hugely more frequent than general spoken English.

Any institutional context develops its own phrases, and the CSA is no exception. Formulaic language, in this sense of words that frequently occur together, are common to any genre of language (see 'Linguistic terms 3 What's 'formulaic language'?'). A crucial question is why these phrases can stand out so starkly. Some of these issues will be addressed in the micro analysis in section 4.5, which makes a close analysis of the 'how' and 'where' of formulaic phrases.

Broad analysis of the 40 cases

3.4.2 'Keywords' – differences from everyday General Practice

A criticism of simulated consultations is that they are not the same as real consultations. In terms of the words used, we can test this by comparing it to language datasets.

- Corpus of GP consultations from the mid-90s – John Skelton, School of Health and Population Sciences, Birmingham = 600,000 words

There were some limitations with Skelton's data: it features no international medical graduate doctors, only one ethnic minority doctor, the patients were all English first language speakers and many of the consultations were lengthy, involving experienced GPs, giving a different feel from the CSA. It was also recorded in the mid-1990s and, arguably, consultations have changed to some degree. However, this is the most recent large data base of real consultations so it provides a useful starting point. We also compared the CSA with the dataset of GP encounters from a project by Celia Roberts (Roberts et al 2005) that recorded consultations in South London practices, including a high number of linguistic minority patients.

- Corpus of GP consultation in Lambeth Practices – from the PLEDGE project – Celia Roberts, King's College London = 90,000 words.

Comparison of CSA data with Skelton's GP consultation data:

When performing corpus analyses of the CSA consultations, the data was found, perhaps quite inevitably, to be closer to everyday GP consultations, in terms of the lexical choices that appear, than to more general spoken English. The table in Appendix B-13 shows all the phrases that occur more often in the CSA than in everyday general practice and it looks very different from the comparison we saw with spoken English. In particular, 'how do you feel' phrases appear, on the surface, to be as frequent in John Skelton's data as they do in the CSA. Phrases, spoken by the doctor, relating to the patient's 'feelings' do actually occur less often in Skelton's corpus but this is not statistically significant, so they do not appear in the keyword comparisons. This suggests then that discussion and phrases around 'feelings' occur almost as often in 'real' GP encounters as in the CSA.

Nevertheless, despite the many similarities, a few important differences show up that make the CSA crucially a little different from consultations in general practice. Firstly, the CSA still has significantly more response tokens than 'real life' GP consultations – all the 'Right Ok' 'Ok Yeah' 'Yeah Right Yeah's etc., which dominate the top 20 keyword clusters. Accounting for why these occur more often in simulations is difficult at this surface level. Perhaps this is a response strategy we use in stressful settings to buy a bit of time between turns or perhaps it signals to the examiner you have acknowledged what the RP has said before moving on, something which need not be vocalised in an everyday GP encounter. Similarly the openings: 'tell me a bit more about', 'how can I help you' occur more often in the CSA. Further down the list are the 'any'/'anything' vague questions which we already saw were highly frequent in the CSA, although the difference here, when comparing to GP consultations, is slightly less pronounced

than it was for spoken English generally (3.4.1).

It seems interesting that, on the flip side, the pronoun 'I' should occur significantly more in John Skelton's GP data, as well as clusters like 'I think' 'I had' etc. than in the CSA – these are the 'negative keywords' (see Appendix Table B-14) – the ones that don't occur as often in the exam. The top 4 negative keyword clusters, i.e. those that occur far more in Skelton's everyday GP consultation data, all related to the first person pronoun. Actions and assertions around the first person 'I' are not used much in the CSA.

These are relatively small differences. Overall, the CSA data is similar to Skelton's corpus of GP consultations, suggesting that much of the talk in the exam, is close, in lexical patterning, to 'real life' practice. The small differences are important to explore further and suggest elements of a special 'linguistic fingerprint' for the CSA. But it is important to note that, overall, the corpus analysis shows, too, similar datasets.

Comparison of CSA data with Lambeth practices data: In comparing the CSA dataset with this more linguistically diverse data from the Lambeth project, a few more interesting differences can be seen. The keyword tables, showing the phrases that are more frequent in the CSA or in the Lambeth GP practice, are given in Appendix Table B-15. The top 20 keyword clusters for the CSA are again dominated by the 'ok'/'and' response tokens. It is interesting though that 'sort of' and 'kind of' occur more often in the CSA than in the Lambeth consultations, a feature that becomes more imperative as we investigate successful and unsuccessful candidates in 4.3.1.3 below. 'Sort of' and 'kind of' were in fact hugely frequent in Skelton's corpus, so not statistically different from the CSA, but this is not the case in these more multilingual consultations in Lambeth where vague 'hedges' do not occur as often. Once again the 'any'/'anything' vague questions appear more frequently in the CSA than the Lambeth data, as well as the 'tell me a bit more about' 'how can I help you' openers, leading us to conclude these are CSA phrases. Likewise, the negative keyword lists (Appendix Table B-16) reveal how infrequently CSA candidates talk in the first person 'I' compared with real GP practices.

Generally, the Skelton corpus and Roberts' Lambeth consultations corroborate that there are a few linguistic features that demonstrate a particular kind of CSA 'fingerprint':

- Response tokens featuring combinations of 'ok/'right'/'and' – more frequent CSA

- Is there 'anything else'/ do you 'have any' questions – more frequent CSA

- 'How can I help you…' / '…tell me a bit more about' – more frequent CSA

- First person pronoun assertions and actions 'I think'/'I will'/'I have' – less frequent CSA than in regular GP practice

There are also important distinctions which crop up, particularly the use of 'kind of' and 'sort of' being more frequent in John Skelton's data than in the Lambeth consultations. These hint towards some linguistic characteristics of multi-lingual and multi-ethnic practices that may not be reflected in the CSA. These, together with the comparisons between CSA diversity tagged

cases and the Lambeth corpus suggest that consultations with linguistic minority patients require and produce some different linguistic features than those assessed in the CSA.

So while the CSA has a particular 'linguistic fingerprint' which makes it different from both the real GP consultation data sets, in some respects, the linguistic norms of the CSA are closer to those of the Skelton data than the Lambeth data because the Skelton data was based on doctors and patients whose first language was English.

3.4.3   Difference between CSA candidates

It becomes clear above that there are certain words/phrases indicative of talk in the CSA. The obvious question to ask is whether everyone uses them in the same way. We compared:

- High scoring candidates with low scoring candidates

- Clear pass cases with clear fail cases

- International medical graduates with UK graduates

Hardly any differences appear (Appendix Table B-17). Some of this can be accounted for by the small word count of the datasets, leading to fewer statistically significant outputs. Nevertheless, for so few differences to appear suggests, at this surface, macro-level, most candidates are talking in the same way, at least in terms of the words they use.

However, there are some differences between candidates. The vague hedges 'sort of' and 'kind of' are statistically more frequent in the successful, high scoring candidates talk than unsuccessful. The cluster 'sort of' again comes up as more frequent in UKGs than IMGs consultations. This ties in with a later finding from the more qualitative work that conversationalising, including the use of vague category markers and hedges, is a feature of successful candidates' talk (see 4.3.2.1). It is interesting that these hedges 'sort of' and 'kind of' also occur so often in John Skelton's GP consultation data, but not in the more multi-lingual Lambeth consultations. We can suggest therefore that these types of vague hedges are a feature associated with local English speakers interacting, and not as prevalent in talk where only one or neither of the speakers falls into this category.

It is perhaps of interest that successful candidates are saying first person pronoun 'I' phrases slightly more often than unsuccessful candidates – perhaps a few doctor centred 'I think' assertions serve well in the CSA and certainly examiners expressed this as a positive feature. Nevertheless, this was not statistically significant and all candidates still say it far less than in everyday GP practices, so 'I' assertions are not a clear cut indicator of high scoring talk.

Overall, the corpus comparison highlights that, there are words and phrases that are very particular to the CSA setting when compared with spoken English and 'everyday' GP consultations in a multilingual setting. This suggests a CSA 'linguistic fingerprint'. Most of the formulaic expressions identified cluster around the interpersonal work of the consultation (phrases using 'feel', response tokens 'OK'/'alright' and questions using 'any'/ 'anything'), which were also identified by examiners (Chapter 5) as the stand out phrases of the CSA. There

are some indications that CSA consultations reflect a more local English – they are marginally more like John Skelton's GP consultation dataset than another dataset of consultations in Lambeth, where GPs were often interacting with patients with less than fluent English. Features such as vague language hedges 'sort of/kind of' used more frequently by successful candidates also point towards a more local, UK variety of English. While these differences are noteworthy and statistically significant, the relatively small size of the CSA dataset (85,000 words) needs to be taken into account. The larger reference datasets of millions of words help to characterize the CSA features with some certainty and it is likely that these CSA patterns would be repeated if we transcribed further cases. However, a much larger body of data would need to transcribed to confirm this tentative finding.

3.5   Analysing interactional features across the 40 cases

In the mapping process described in 2.5, we annotated transcripts not only for their broad phases, but also smaller interactional features, the frequencies for which are given in Appendix Table B-18.

Role-player sink/save – moments where the role-player appears to help or hinder the candidate in the success of the case or in creating mutual alignment. It is clear from the figures in Appendix Table B-18 that there is no consistent pattern in this overt role-player behaviour that would indicate any correlation to candidates' marks. In the more microanalysis in Chapter 4 below, very small differences in role-player behaviour can change the focus a candidate gives to the consultation and or cause difficulties or uncomfortable moments. So, although very local difficulties can occur, overt role-player sink or saves did not lead to success or failure in the exam.

Alignment – speakers bringing their talk into line with each other to reach a common goal and sustain the flow of the interaction and expressions of overt understanding, agreement or compliance (`right i can understand why you would be worried`) (Detailed further in Section 4.3).

Moments of expressed alignment occurred throughout the transcribed dataset, and while they might be thought of as a generally positive indicator, they only occurred marginally more in the 'passing' cases than in the failing. Outright failing cases had an average of 2.9 of such moments per case, ranging from 0-9 instances, whereas passing cases had an average of 4.4, ranging from 2-11 instances. 'Borderline' cases, either passing or failing had around 3-3.2 per case. Overall such interactional features can therefore be seen as a positive indicator of performance, but only by a small margin. Moments of expressed alignment were also a feature that showed up slightly more in UK graduates' (average 3.4 per case) consultations than in IMGs' (2.7 per case), regardless of how well the candidate did overall. Cultural/linguistic factors may therefore play out in how successfully these moments are performed and perceived in the simulated exam setting.

Broad analysis of the 40 cases

Misalignment/Misunderstanding – moments of overt misunderstanding, communicative 'trouble' between candidate and actor, expressions of disagreement (Detailed in Section 4.4). These showed a pattern between poor and high performing candidates: outright failing cases had an average of 5.8 moments of misalignment per consultation – ranging between 3 and 15 instances in each case. This compares to an average 3.6 instances in passing cases and 4 in 'borderline' passing, ranging between 0-8 instances. Although the difference is less stark, misunderstandings and misalignments were also a feature more frequent overall in the IMGs consultations (4.5 misunderstandings per consultation) compared to UKGs (average 3.3 misunderstandings per case). The reasons for this slight difference are explored in Chapter 4. Nevertheless, almost none of the consultations were entirely free of misalignments – many successful candidates, from the UK or overseas, experience some moments of communicative trouble. Importantly, many are able to repair these difficulties, and the success of this recovery work forms an important part of the discussion and training material from the project. (So, while attempts at repair when the overall timing and structure are weak are difficult for all candidates to recover from – see 3.2.3 above – culturally appropriate repair from misalignments and misunderstandings can still lead to passing the case.)

Exam Modelling – features indicative of a CSA 'exam model', encompassing 2 key ideas:

- Lexical phrases oriented to the patient-centred model that stand out in 3.4 and in examiner feedback as being more frequent in this context than in everyday GP consultations.

- Any utterances overtly addressed to the exam situation – things must be articulated in an exam setting to demonstrate knowledge and understanding of the situation to the examiner which might not need to be articulated to the same extent in real, day-to-day GP consultations e.g. statements about what the doctor will be looking for in a physical examination, prefacing statements, explicit questions and statements that display patient-centredness etc.

Elements of exam-modelled strategies and phrases were features of all candidates' talk, although higher performing candidates actually use them more. For lower scoring candidates, an average of 6.7 exam modelled utterances were identified per case, compared to an average 12.1 for higher scoring candidates. Borderline cases averaged 8 utterances identified as exam-modelling. Overall then, candidates who used more exam-modelled phrases tended to do better. This shows that there are expectations about how the exam should be performed, and that there are certain questions and expressions which have to be included. However, many of these exam-modelled phrases are also seen as formulaic when played back to examiners (see 4.5 below) and it is the 'how' and 'where' of their use which makes the difference. So, there is a paradox. Successful candidates do a lot of exam modelling but this is also frequently seen as formulaic or causes misalignment in the performance of lower rated candidates. This suggests that subtle differences in manner can make a difference to candidate outcomes and that being taught 'phrases' for the exam may be a problem not a solution (see Chapter 4).

Summary – Some interesting patterns emerge from these annotated interactional moments that merit closer analysis. In particular, given that annotated moments of alignment, misunderstanding/misalignment and exam-modelling showed a relationship to candidates' marks, as well as to features identified by examiners in feedback sessions, it seemed important to analyse them at a more detailed and local interaction level. For while differences are shown, they are often marginal, and many candidates are doing the same sorts of things. Without a closer analysis, it is difficult to see how they may be working to help or hinder the candidate.

Chapter 4.
Micro-level analysis of the 40 cases

4.1   Summary

Throughout the chapter, we observe how small differences at a local interactional level can have large consequences in how the consultation plays out. It looks at how alignment, misunderstandings and misalignment, exam modelling and formulaic language, and explanations occur in the local, turn-by-turn talk. These small sequences can impact significantly on the subsequent talk and, cumulatively, can lead to a global impression of how a candidate communicates (see example 4 – 2). This goes much further than the more general analysis of Chapter 3 in explaining why some candidates came across as poor during our examiner video feedback sessions and indeed in the exam marks.

- 'Alignment' and 'misalignment' are proposed as more useful than terms such as 'empathy' or 'warmth' in judging how the interaction is managed in this simulated setting (see Linguistic terms 4, and appendix E-1). This is not to say empathy is unimportant to medical consultations, but rather an acknowledgment that what can be judged in the CSA, as a simulated setting of clinical interaction, is communicative behaviour of candidate and role-player, not inner emotional experience. The term 'alignment' gets away from the 'doubly subjective' nature of fathoming how the RP might be feeling as a fictitious patient. Nevertheless it is important to emphasise that no wholly 'objective' means of judging interpersonal skills can be achieved – how a person talks is necessarily evaluated against prevailing norms and 'how I would do it' (Yeates et al 2013). The difficulty in deciding how to mobilise and measure this subjectivity and how much linguistic tolerance can be built in is discussed in Chapter 6. 'Alignment' also poses a useful concept for the learning materials, as a way of thinking about the turn-by-turn interaction and not simply a recipe like set of 'empathy' phrases.

- The line between achieving alignment and misalignment is a fine one. Many candidates in fact use the same/similar strategies and exam phrases but with differing success. Many of the strategies to achieve alignment can backfire and cause misalignment. The particular location within a given sequence of talk has a profound effect on success or failure e.g. rapid topic shifts or locating 'empathy' phrases briefly within biomedical sequences can appear tokenistic and insincere to examiners. There are several strategies that successful candidates use to bleach the 'formulaic' from exam phrases and sound more sincere: customising the wording a little, using formulaic words such as 'feel' in other contexts so they lose their recipe quality, using expressive non-verbal features. Successful candidates also use conversationalising strategies, such as vague language and softeners, to create a casual conversation style within this institutional interaction.

- Responses of the role-player can mark out 'alignment' strategies as successful or not. Alignment is a joint production and the RP's responses are a powerful channel in how a candidate comes across. Roleplayers flag up misunderstandings in the CSA more

than patients tend to in real life GP encounters and candidates tend to be on the back foot apologising. While RPs' overt 'sinking or saving' behaviour showed no systematic effect on candidates marks (see Chapter 3), their more powerful position in the interaction can constantly be seen, more subtly, in the localised turn by turn responses. This power is inevitable in assessment simulations but cannot be underestimated in terms of the communicative differences it causes in the CSA compared with everyday GP consultations.

- Often small and subtle cues in RPs' talk need to be picked up through accurate listening and inferencing. Missing these can lead to misalignment and sometimes even missing the key nub of the case. Mastering this local interactional feature of inferencing is therefore crucial to figuring out how to direct the more global structure of the case and has implications for our e-learning material.

- Virtually all candidates experience moments of misunderstanding and/or misalignment, but poorer performing candidates have slightly more. Misunderstandings can occur for various reasons, but attempting to repair trouble can be difficult. Where the trouble was repaired swiftly and located appropriately it can be rapidly defused. In all other cases, it is difficult to recover from mistakes or missed opportunities later in the consultation without losing at least some marks. Some misunderstandings are directly caused by the simulated setting itself. The very high number of disfluencies and false starts might also be a factor of exam nervousness. Disfluencies and mistakes occur in all talk but we show that they are more likely in the exam context, raising the question of whether a position of greater communicative tolerance could be adopted.

- 'Metacommunication', or talk above the literal talk (see Linguistic terms 6), is hugely important. Much misalignment results from a lack of commentary by the candidate on what they are doing and what stance they are taking. Special attention should be paid in CSA preparation to explanations of why a physical exam is being undertaken and what is routine or, possibly, concerning. Metacommunication is also a feature of all every day and professional talk (see Linguistic terms 6). For example, it is a strong feature of explanations, with the candidate commentating on how new information links to the preceding talk. Showing explicit awareness of exam requirements is a feature of many face to face assessments but, in addition, candidates in the CSA are rated highly if they 'talk aloud' about their actions and their perspectives.

- Explanations in the CSA fall into four broad categories (section 4.4), each requiring a somewhat different focus. All explanations, though, require extended talk, balancing the linking of complex information and ideas with how that information relates to the particular patient, delivered in a 'story-like' structure. All candidates in fact display some of the features of successful explanations, but a moment by moment lack of coherence and explicit tying of ideas together can undermine these successful strategies.

Micro-level analysis of the 40 cases

- Amongst failing candidates, a range of communicative differences, mostly subtle and difficult to analyse in on-going talk, can lead to explanations appearing to lack coherence and lead to misalignment. These communicative differences are analysed in relation to the English spoken in the Indian sub-continent, where they are used systematically, and from where the largest group of IMG candidates come. In sum, they lead to difficulties in tracing the line of argument and processing how phrases and clauses link together.

- In noting how successful candidates could bleach out the 'formulaicness' of their exam strategies, it is important to consider that it is challenging for candidates whose English is influenced by other languages (and this is described for speakers of Indian English in section 4.5.3) not to sound formulaic since their use of intonation and other prosodic features may be, at least partly, used differently from local English speakers to express affect. People's use of language is of course not fixed and many candidates from the sub-continent will be influenced by local English, displaying some or none of the features described here.

When listened to in the 'real-time' setting of the exam, all these features are processed by examiners quickly and largely below the level of consciousness in building up an impression of candidates' talk. It would, of course, never be possible to think in this detail while marking a candidate. We hope though to show how 'putting the microscope' on talk, reveals the very detailed level at which CSA interactions can be successful or not. We also propose some means of conceptualising interaction at this localised level, both for helping candidates think about their consultations in preparation for the exam, and for the exam configuration itself.

4.2   Introduction

That conversation is achieved through moment-by-moment talk is clear, but the complexities of managing this become hidden. It is a skill so well learned that we have conversations all day, every day, without thinking about the intricate, real-time task we are performing. We implicitly know how to manage the messiness of talk, as well as tailor our talk to particular settings and communities. These hidden complexities can be brought to the surface in figuring out how tasks are achieved or where things go wrong. In an institutional and assessed setting, these usually unconscious features of interaction become more demanding, with significant consequences.

Through the mapping annotations outlined in 2.5, the research team systematically identified micro-interactional features that occurred consistently across cases. General findings on these were presented in 3.5, where some patterns emerged related to candidates' success or failure. This chapter looks at these micro level moments in their context of use, specifically:

1. Alignment between candidate and RP

2. Misunderstanding, misalignment and repair

3.	Exam modelling and formulaic talk

4.	Explanation phases and their particular micro-interactional features

These were the interactional features which showed the strongest correlations to weak or strong CSA performance in 3.5 and this is the justification for concentrating on them here. Explanation phases were also identified in 3.3 as among the most complex communicative tasks to achieve, with a heavy load on the candidate to do the majority of the talking, and are therefore selected for closer examination in understanding candidate success or failure. It is at this micro-level that some account can be made as to why the quantitative analysis in Chapter 3 could not always show a difference between the communicative styles of successful and unsuccessful candidates. It begins to highlight how small features of interaction can have large consequences, particularly where the features occur accumulatively over the course of a case. This ties in with some of the findings of the examiner feedback sessions (Chapter 5), where small interactional features impacted on the appraisal of candidates. Analysing these fully contextualised, real instances of talk is crucial for developing learning materials that are relevant and accessible to trainees.

The mapping of alignment, misalignment and misunderstanding moments was based on widely used sociolinguistic methodology (Gumperz 1982, Coupland, Giles and Wiemann 1991, Hinnenkamp 2003, Roberts et al 2005) and each example was identified and agreed by SA and CR.

As with any qualitative study, requirements of space make it impossible to report every example and only a handful of illustrative cases appear. These 'telling cases', while they occur within their own contexts, help to illuminate wider patterns and theoretical relationships. Should the reader be interested, further examples, which formed part of the working, are referred to in Appendix C.

4.3   Alignment

A central area of concern in the CSA is to assess if a candidate is patient-centred. This requirement permeates all domains, but crucially coalesces around interpersonal skills. Sustaining social relationships is a highly indicative feature of talk in the CSA. The corpus linguistic analysis in 3.3 demonstrated the most frequent CSA phrases were oriented towards the relationship work of interaction. This tricky area of sustaining social relationships through talk is discussed throughout the chapter, but we begin by thinking about an analytic vocabulary for this context, for which we suggest the concept of 'alignment' (Stokes and Hewitt 1976) with the patient. Alignment is not such a charged or idealistic term in this role-played setting as 'rapport' or 'empathy', or showing 'connection', used in the classical consulting models. It deals with conduct and that is all the CSA can do as a simulated test (Miller 1990).

Linguistic terms 4

A vocabulary for social interaction in the CSA

The interpersonal skills domain is characterised by descriptors requiring 'empathy', 'rapport' and 'connection' (see Chapter 6 for analysis). While 'empathy' might be desirable and part of the important caring role of primary healthcare practitioners, there are several reasons why 'alignment' (Stokes and Hewitt 1976), is a more useful substitute here;

(i) 'Rapport', 'empathy' and 'warmth' are tricky if not impossible to judge. Although we all may have an intuitive, commonsense understanding, ultimately these are inner emotional experiences of the interactants that are hard to observe from outside. It is doubly subjective: observers can only guess at the individual, inner reactions of listeners.

(ii) Empathy, as an emotional response to another person, becomes complicated in a simulated, acted setting where there are very different emotional reactions at work.

(iii) Interpersonal skills are assessed in CSA descriptors as created solely by the talk of the candidate alone. But interaction is a complex, joint production, like 'climbing a tree that climbs back again' (Erickson 1986: 316). This joint work is important to evidence.

(iv) 'Alignment' moves away from a psychological way of assessing consulting skills to a social/behavioural way – looking at evidenced professional performance rather than inner feelings. This functional evaluation is useful in assessing simulated encounters.

(v) 'Alignment' describes aspects of the consultation which deal with its overall management (managing the interaction and its speed and rhythm and making appropriate inferences from the other's talk), as well as explicit strategies to show understanding and, where appropriate, agreement.

(vi) The institutional and professional constraints of the GP's role mean that not every ethical stance expressed by the RP can be supported with 'empathy'. However, the doctor is expected to maintain the flow of interaction and a level of mutual understanding. The term 'alignment' allows for the tension between sustaining social relations but not necessarily fully supporting/endorsing the RP's stance.

Features of alignment can be usefully discussed in terms of whether they achieve 'Interactional' or 'affiliative alignment'. Interactional alignment describes how speakers bring their talk into line with each other in trying to reach a common goal and sustain the flow of interaction. It indicates mutual understanding of the talking task, even if not necessarily ideological agreement between speakers. It assumes that each side is giving the other the opportunity to speak without being impeded. This is crucially dependent on the task at hand. For example, in 3.3 we saw the explanation phase required understanding from both participants that the GP should do the majority of the talking here. Much of this interactional alignment occurs below the level of consciousness, such as speakers converging towards using the same expressions or mutually agreeing on topic shifts. Interactional alignment also encompasses problematic,

disfluent moments that most talk encounters, such as dealing with uncertainty or unclear meanings. Alignment work here requires getting things back on track, although as section 4.4 on misalignment will explore, this is not always achieved.

Unlike the more unconscious aspects of interactional alignment, affiliative alignment is identified when either speaker expresses agreement or explicit orientation to the concerns, feelings or expectations of the other. These include: eliciting patients' stance on an issue, establishing knowledge and expectations, conveying tokens of mutual understanding or agreement. All of these are features of CSA talk. An example of this idea is in a highly agonistic case (see Linguistic terms 2 'Agonistic dialogue') when a successful candidate asks the role-player how he will 'disguise' any medical effects of his requested secret vasectomy (see Example 4-16, Section 4.4.4.2). 'Disguise' is a potentially disapproving description, a 'discrepant word' that misaligns with the patient's stated view of his actions. The candidate rapidly deals with this by highlighting the inappropriacy; `I'm sorry, disguise is maybe the wrong word to use but (.) you know what I mean`. While she might not agree with his decision, explicitly expressing a level of alignment enables the complex discussion of the case. Attending to this kind of alignment, or the positive 'face' of the patient, becomes especially complex in situations with asymmetric power, although as Linguistic terms 5 in the box above discusses, the interactional power differentials in the CSA are not clear cut.

> ## Linguistic terms 5
> ## Power in interaction
>
> Power is an important topic of concern in linguistics and the means by which speakers exert power in interaction have been variously argued (such as interruptions, questioning strategies, highlighting mistakes). The GP consultation is typically characterised by asymmetric power, with the GP holding the ability to provide treatment. Power differentials by their nature mean speakers don't quite align, although a sense of alignment can be talked into being (Drew 1991: 21). In the simulation, that asymmetry is reversed, with the RP already aware of the correct management. This engagement framework imports another level of complexity to the setting.
>
> We show indications that the RP holds a more interactionally dominant position (see 3.2 and 4.4.1). Achieving alignment in this setting is a complex, hybrid task, mimicking the dialogue of a real consultation but within the additional requirements of this altered power structure.

4.3.1   Interactional alignment

4.3.1.1 *Figuring out the nub of the case*

Interactional alignment work happens right from the start of the consultation and in this standardised setting the interactional flow has an especially fixed pattern, with a standardised opening line from the RP, usually followed by a highly formulaic phrase like 'could you tell me a bit more about...', (see 3.4 and Appendix Example C-1). Further putting a microscope on these opening lines shows how small and subtle differences even at this most standardised of

sequences affect interactional alignment between the speakers, as well as ultimately aligning the candidate to the key nub of the case. These consistent opening lines begin to demonstrate subtle difference in communicative choices. The following two extracts (Example 4-1 and Example 4-2) concern two different candidates, but feature the same RP and scenario. They relate to a patient asking about glucosamine for osteoarthritis. The candidate is charged with the task of deciding how to deal with the patient's request for a drug on prescription that is not clinically proven to be helpful.

Example 4-1 – Opening from '*MS 1*', IMG, Male, FAIL

```
36 CAN: how can I help↘
37      (0.7)
38 RPL: er my daughter was on the internet doctor and she was (0.5)
39      she told me about (0.6) something called (.) glucosamine↘
40 CAN: okay→
41 RPL: er (0.7)
42      for my joints→
43      (0.8)
44      er I wonder if you (0.3) thought it might help
45 CAN: okay (.) so in in
46      (1.0)
47      I I can see that you've got arthritis in your knees→
48      ⌈is that why⌉ you you thought→
49 RPL: ⌊mmm⌋
50 CAN: that it mi⌈ght help or→⌉
51 RPL:           ⌊that's that's⌋ what's been tr- troublesome yes↘
52 CAN: ok and it→
53      (0.8)
54 CAN: it (0.5) you just said it was troublesome how troublesome is it→
```

Example 4-2 – Opening from '*MS 2*', UKG, Female, PASS

```
36 CAN: how can I help↗
37      (1.5)
38 RPL: urgh
39      (1.6)
40      my daughter was on the internet doctor (0.2) and she→ (0.8)
41      told me about (0.4) something called (0.9) glucosamine↘
42 CAN: okay↗
43      (0.6)
44 RPL: I wondered (1.0) whether you thought it might do (0.4)
45      any good→
46 CAN: okay↘
47      (0.6)
48      what do you know about glucosamine↘
49 RPL: well she gave me a right I I've I've spoken about it to to
50      friends in the past≈
51 CAN: ≈mmm
```

The openings pass off almost identically. Both candidates initiate the consultation by eliciting the reason for the visit and the RP describes a family member's online search for medication. What follows, however, is that the two candidates take up different topics as the focus, sending the consultations down divergent paths. In Example 4-2, the candidate (UKG, PASS) addresses

the inferred request for glucosamine, while the candidate in Example 4-1 (IMG, FAIL) homes in on the patient's arthritis. When looking at how the consultations unfold, we see that the topic of glucosamine is not reintroduced in the first case until much later in the consultation. This unsuccessful candidate conducts a slightly late data gathering phase (see Figure 6) where he must revisit the glucosamine issue after 'half time'.
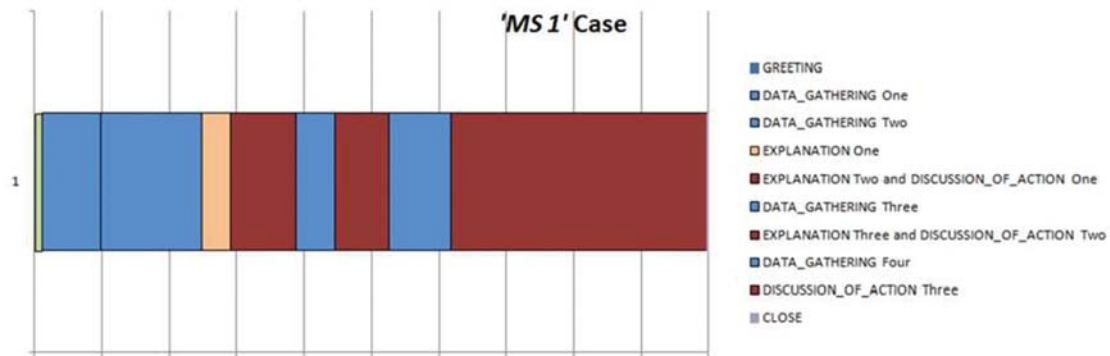
Figure 6 – Unsuccessful candidate from Example 4-1
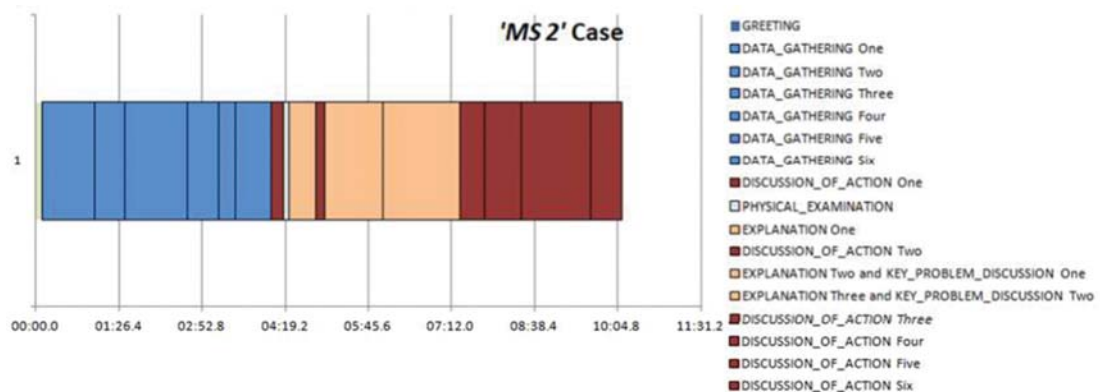
Figure 7 – Successful candidate from Example 4-2

It is arguable that the slight difference in the RP's opening gambit in Example 4-1, in which, unlike Example 4-2, the patient mentions 'joints', might lead the unsuccessful candidate to take a slightly different route in asking about arthritis. In any event, what the examples demonstrate is that small differences in how a case is responded to early on can potentially lead to a data-gathering structure and even a whole case unfolding in different directions. Such trouble, as we saw in 3.2.3, can be difficult to recover from in the 10 minutes. Small differences can have large consequences for the interactional flow and successful alignment with the nub of the RP's wants. Even highly standardised interactions then experience these important micro-level divergences. This has implications for e-learning material, in terms of inferencing and how to pick up on the nub of the case or repair misdirection, perhaps a useful skill for all consulting.

4.3.1.2 Managing interactional flow and inferencing

Small difficulties in interactional alignment may also cause a consultation to appear more generally 'clunky', as many examiners described in feedback. The CSA can test how candidates manage difficult patients, meaning that role-players may be expected to give unexpected or discrepant answers that challenge the flow of interaction. On occasions, the discrepant answer is clearly signalled by the role-player, as in Appendix Example C-2 from 'KM 1', a role-played unresponsive teenager. But RPs also give more ambiguous answers, from which the candidate must infer that the patient is not that compliant:

```
Example 4-3 – 'RS' Case, IMG, Female, PASS
364 CAN: what's exactly the problem and helps you to deal with it and
365 RPL: right↘
366 CAN: helps you find ways to deal with deal with it
367      so how do you feel about that↘
368 RPL: erm (0.4) yeah yeah could do↗
369      [er:] is that referral is it↗
370 CAN: |mhm|
373 CAN: well we ha- (0.2) er there is er:
374      we have got a (0.4) counsellor in in our surgery↗
375      (0.5)
376 CAN: I could speak to him and see if if he can arrange the (.)
377      a quicker appointment
```

In Example 4-3, at line 368 the roleplayer produces a less than enthusiastic response. The hesitation markers, a pause, the phrase 'could do' tempering the outright acceptance of the 'yeah yeah', display that even though the proposal is met with acceptance, it is only partial and reluctant. So this successful IMG candidate then enforces the proposed course of action, 373-7, attempting to make it more attractive and stressing that a counsellor can be provided quickly. Being able to infer from the RP's behaviour that he may not be compliant, and quickly imagining that there is an in-house counsellor, draws on both interactional alignment skills (accurate inferencing of subtle verbal cues and other markers) and on skills of simulation (as well as problem solving and teamworking skills). These examples show that interactional alignment work and the larger task of agreeing a plan, crucially depend upon these small, micro-level skills of quickly making inferences from the talk and linguistic and bodily cues of the RP.

4.3.1.3 Conversationalising the consultation 1

The corpus studies in 3.4 showed the frequency of small tokens of language that act as 'glue' between utterances. Some of the most frequent examples related to response tokens in varied forms: 'right ok and...' 'right ok ok....'

```
Example 4-4 – 'RS' case, IMG, Female, PASS
CAN: right ok and has she had (.) partners        before
RPL: yeah (0.2) she's she's had a few
CAN: mn right ok right and er (1.1) you s- you
     mentioned that this has been going on (.)
     since you are eighteen
RPL: yeah
```

'ok and…'. Such tokens signal the candidate's acknowledgement of the RP's turn as well as a transition to the next question or topic, particularly during data-gathering. These are often called 'discourse markers' (Schiffrin 1987) and function to enhance the mechanics of interaction, as a kind of punctuation to make conversation fluent (Carter and McCarthy 1995).

The extraordinary high frequency of these response markers in the CSA is perhaps due to an increased need to signal links in the discourse in a simulated performance, as well as demonstrate what was widely called 'active listening' by examiners. Markers of casual conversation have in the past have been almost stigmatised as ungrammatical (Miskovic-Lukovic 2009: 602). But contemporary workplace cultures have seen a shift towards casualisation as desirable, and this would seem to be borne out by the CSA data which produce a kind of casual conversation/institutional hybrid. These very small level, conversational style features are unlikely, in a single instance, to mark out what is seen as successful 'rapport' by an examiner. However, used cumulatively and alongside other features of alignment, they can build an impression of a flowing interaction.

4.3.2   Affiliative alignment

As well as the subtle interactional management of turn-by-turn talk, more explicit nods to the patient centred paradigm can be seen throughout the consultations. They attempt to demonstrate what we termed 'affiliative alignment '(4.1) and unsurprisingly, they are hugely frequent in the CSA.

4.3.2.1   Conversationalising the consultation 2

A range of other strategies, such as vague language, softening/mitigating grammar, and indirectness, work to make many consultations more like British conversation, conveying the stance of the speaker by the content of their words (Aijmer 2002: 39, 48). Here, in Example 4-5, a vague 'sort of' at 421 is used to soften the potentially threatening 'calm yourself down', as well as the distancing 'some people' (line 423) rather than referring directly to the patient. Further examples of

```
Example 4-5 – 'KB', UKG (BME), Male, PASS
413 CAN: first of all you can try and do some act-
414      exercise kind of to try and help
415 CAN: release your natural endorphins
416 RPL: uh-[ huh]
417 CAN:    [and this is] something that you can try and
      do in the evening
418 CAN: or when you get little bit time to yourself
419 CAN: [cos often] that's quite sort of helpful
420 RPL: [yeah]
421 CAN: to sort of calm yourself down
422 CAN: (0.8) er the other thing that
423      some people f- might find helpful
424 CAN: is um
425      you can actually do some relaxation techniques
426 CAN: [the things] like er
427 RPL: [yeah]
428 CAN: yoga or gym classes
429 CAN: these kind of things that
430 CAN: i suppose in similar kind of thing to exercising
431 RPL: m[ m]
```

vague language can be found in Appendix Example C-3, particularly the frequent 'anything' questions, which encompass a broad range of category information ('does it feel hot…or anything like that?') but also index the tenor of casual conversation. The vague markers 'kind

of' and 'sort of' were identified in 3.4 as more frequent in British English consultations and UKG candidates talk. Such deformalising items index a more equal relationship between doctor and patient, particularly useful in sensitive contexts with power differentials. Everyday 'social chat' leaks into all kind of institutional encounters, but knowing how to do this within the institutional architecture requires some subtle linguistic knowledge. Of course, the power differential is reversed in this simulated context (see Linguistic terms 5), but these conversational softeners still mimic British English consultations and, as argued in 4.5, work to make exam modelled phrases sound less formulaic. Again, as a small interactional feature these accumulate to index the 'casual conversation hybrid' style of British consultations and are an important component of success in the CSA.

4.3.2.2   Establishing ICE and 'metacommunicating'

This section illustrates Neighbour's (1987) classic 'ideas, concerns and expectations' formula. Establishing pre-existing knowledge and expectations is an act itself intended to convey alignment. It is frequently attempted in the CSA in a way that explicitly signals this motive, a type of exam-oriented metacommunication, or a commentary on your talk. This term is explicated further in Linguistic terms 6. In Example 4-6 below (lines 124-6), '`then I can build upon your knowledge`', clearly metacommunicates to the examiner the candidate's motives for requesting prior knowledge. The desirable response to ICE questions is for the RP to offer an account of their knowledge, as in lines 128-136. Many successful candidates use this strategy throughout their cases, with long responses from the role-player, as for example with the successful '*MS 2*' examples (Appendix Example C-4). However, when

Linguistic terms 6 Metacommunication

This is an important idea we'll come back to throughout the chapter. Metacommunication is the talk we do 'about talk'. Most people have encountered the idea of 'signposting'- letting the patient know where the consultation is going and what's coming next.

```
*CAN:    it's a very effective procedure but it's •
*CAN:    something with quite permanent ⌈consequences⌉ •
*ACT:                               ⌊mmm mmm⌋ •
*CAN:    which is something that we will go |over •
```

But more subtle metacommunication goes on all the time to give a message about and 'above' the literal message, indicating how we want our talk to be received. When a GP candidate adds '…but it's nothing for us to worry about at the moment', she is signalling how a piece of medical information she has just given should be interpreted, as well as that she is attentive to the relationship with the patient and their concerns. Metacommunication is especially useful in a setting with an overhearing party, such as the CSA. It shows the examiner the structure you are working to, why you are doing something and how attentive you are to patient-centredness. This is a feature of many successful candidates' ways of talking in the CSA.

the RP does not respond with a long description, and especially when they give a short negative (as was the case with *'KM 1'*, 4.3.1.2 above), the request can stick out more awkwardly. These more discrepant, negative responses from the RP tend to cause more hesitation and trouble in interaction from the candidate. In another *'JA'* case in Example 4-7, the RP responds to the question on whether she knows about cystic fibrosis fairly minimally in line 242 'yes yes I do'. Without a more fully elaborated background knowledge to comment on, the candidate stumbles a little at lines 243-6 'ok right so (1.8) um'. When asked how much she knows about chromosomes, the RP's discrepant, 'I wouldn't know where to start' again results in a hesitant 'sure ok hhh so i mean um' (255-6) from the candidate, who, crucially, does not metacommunicate any of his intentions here. Using these open ended strategies to request prior understanding is done by most candidates, but the responses of RPs do not always evidence 'alignment' – quite the opposite if an unhelpful response is offered. The appearance of alignment is therefore co-produced at these local levels, where the RP can behave in subtly different ways. Learning ICE questions, which all seem to know, would not therefore seem to be as important as knowing how to successfully embed them in the interaction and 'metacommunicate'.

```
Example 4-6 – 'JA 3', IMG, Male, PASS
124 CAN: that's all right how much do you
         know
125 CAN: ⁇tell me⁇ (.) what do you know about
         cystic fibrosis
126 CAN: then i can build up on your
         knowledge
127 PPP: (0.4)
128 RPL: er we- i: know that's it's erm (0.5)
129 RPL: a disease which causes (.) problems
         for (.) breathing
130 RPL: |and|
131 CAN: |right| that's right |xx|
132 RPL:                       |and| the
         bowels
133 RPL: (0.5)
134 RPL: erm that people with cystic fibrosis
135 RPL: often don't live as long (0.4)
136 RPL: a life
137 PPP: (1.4)
138 CAN: mhm
```

```
Example 4-7 – 'JA 1', UKG (BME), Male, FAIL
237 CAN: and um in in
238 CAN: terms of cystic fibrosis do you
         know how it affects someone
239 PPP: (1.1)
240 RPL: er about the disease
241 CAN: |yeah|
242 RPL: |yes| yes i do
243 CAN: ok right
244 CAN: so
245 PPP: (1.8)
246 CAN: um (.)
247 CAN: do you know much about chromosomes
248 PPP: (1.3)
249 RPL: not scientifically no I've heard
         of them |but|
250 CAN:          |ok|
251 CAN: would you mind telling me how much
         you know
252 CAN: that's going to
253 RPL: i i wouldn't know where to start
254 RPL: |really|
255 CAN: |sure| ok ·hhh
256 CAN: so i mean um
```

4.3.2.3   *Life-world questions and the joint production of 'empathy'*

As part of a patient-centred philosophy on the holistic effect of illness, almost every consultation involved questioning the patient on their day-to-day life and effects of health problems. Given that candidates routinely ask these questions, there is, on the surface, little to differentiate high and poor performing candidates. Differences do occur in how candidates respond to role-players' answers or locate these questions. A failing UKG candidate in Appendix Example C-5, asks 'how are you doing at home?', to which the patient replies with a lament about the things he can no

longer do. The candidate's response 'OK fine' and next question do not seem to adequately respond to this cue. Similarly the rapid topic shift in *'JA 1'* (Appendix Example C-6) away from the discussion of her niece with cystic fibrosis would seem to contribute to his poorer performance in interpersonal skills. The localised sequence is important then. Often weaker candidates shift rapidly between psychosocial questions and other medical or QOF (Quality Outcomes Framework) type questions, explored further in 4.5.2 on 'formulaic' language. This has strong implications for the way such patient-centred strategies are presented in training material.

The preferred response after asking these, or after the RP has volunteered information about difficulties, is to verbally express 'empathy' or understanding for the patient in some way. This requirement was frequently expressed in examiner feedback (e.g.: 'the first thing you'd say "I'm sorry to hear that, what happened". It's just a human response and she didn't do that' Examiner Feedback 2). Not explicitly doing what is called 'empathy' at all does not appear to be an option as every candidate, bar one, attempts expressing it at some point in the consultation. Opening lines from two apparently identical interactions with the same RP (Example 4-8 and Example 4-9) show how small interactional and localised differences may affect the joint production of alignment and perception of a candidate.

The sequences are almost identical. In line 1, the candidate elicits the reason for the visit, which the RP provides (lines 2-5), describing a letter he has received about his father's death. In lines 6, however, the candidates respond slightly differently.

The earliest point that the candidate could express 'empathy' is immediately following the receipt of the information on the '`father's death`'. In the first opening, Example 4-8, the candidate has acknowledged the RP with head nods and a minimal receipt 'okay' (line 5). This puts the onus on the RP to expand the initial account, but he further marks an end to his account with a minimal '`mmm`' (line 7). This provides the candidate with the next turn slot but, again, rather than producing any display of alignment, she produces another sequence-closing token ('`okay`') and focuses on the letter (line 8). This candidate has identified the chief task of the case as the haemochromatosis and genetic risk explanation, attempting to get into this early on by starting to establish the patient's prior knowledge. While she's not entirely wrong, as this is the main work of the case, the failure to take these small opportunities for expressing token empathy responses would seem to have a cumulative effect on her marks – losing a mark on interpersonal skills. In the second examination, Example 4-9, the candidate does produce an explicit 'empathy' phrase (line 6), and indeed does this at the earliest possible point, in overlap with '`last year`'. The candidate follows this with an almost acted account of being aware of this back-story, referring to the notes, '`i did hear`', a good example of imaginative acting skills. The role-player accepts the display of compassion, by downplaying the relevance of the loss (line 8), thereby closing the sequence – they can then get on with the similar work of the first candidate in discussing haemochromatosis. However, this has been more smoothly transitioned. Importantly then, this small early interjection with an 'empathy' token allows for the RP to orient to the candidate as having displayed appropriate alignment, in the right place

```
Example 4-8 – 'MH1', UKG (White), Female, FAIL
1 CAN: what can I do for you
2 RPL: erm well I had a letter from the hospital
3      erh asking me to come and see you
4      in connection with my father's death: (0.2)
5      last year
6 CAN: okay
7 RPL: mmm
8 CAN: okay (.) did the letter tell you anymore
```

```
Example 4-9 – 'MH 2', UKG (White), Male, PASS
1 CAN: what can we do for you today
2 RPL: er well I had a letter actually from the hospital
3      asking me to come and see you
4      in connection with my father's death
5      |last year|
6 CAN: |sure sorry| about that i did hear
7          (0.7)
8 RPL: yeah well as i- you know it was a year ago [now so]
9 CAN:                             |yeah|
10 RPL: yes
11 CAN: sure ok
12 CAN: um and i mean i take it y-
13 CAN: you want to kind of know a bit more about about things
14 CAN: is that right
15 RPL: yes absolutely i mean i i i
16 RPL: i know from the letter obviously it's to do with
17 PPP: (0.7)
18 RPL: um
19 PPP: (0.8)
20 RPL: you know the haemochromatosis that
```

and right manner – quickly in a relatively low tone and backed up with the imaginary detail 'I did hear', before then smoothly moving on.

Small differences played out at this local level are where 'empathy tokens' successfully align, or not. An analysis of another set of opening sequences from the 'JA' cases, one passing (an UKG ethnic minority candidate) and one failing (a UKG ethnic minority candidate, with a feedback statement that the candidate fails to 'show sensitivity to the patient's feelings), can be seen in Appendix Example C-7. Again the ability to make subtle inferences from the RP's turns is needed in order to provide the required 'empathy phrases' when talking about the RP's imagined niece.

Expressing understanding was used often strategically to precede a discussion of institutional tensions, such as why a patient could not have glucosamine on the NHS in 'MS 2' (see Appendix Example C-8). These 'empathy' phrases can be a useful tool for getting into the institutional work of the consultation, while still balancing a professional level of alignment with the RP, an essential 'rule of the game'. However, empathy tokens also seemed to be a constant source of criticism where perceived as too brief and formulaic, explored further in Formulaic Language (4.5), so are problematic to achieve. This is important to address pragmatically in learning materials.

4.3.2.4   Commending the 'good' patient

A similar alignment strategy that appears in lots of cases, successful and unsuccessful, is commending the patient. This can be for good health decisions (Appendix Example C-10) or commending the patient's knowledge (Appendix Example C-11). Candidates strategically use this approval to preface a more institutional, less aligned stance. For example, the doctor commends '*MC 2*' for her son's management of schizophrenia, before immediately saying a medication cannot be prescribed (Appendix Example C-12), or '*TJ*'s secret vasectomy, where the doctor approves it's 'your decision to make' before suggesting alternatives (Appendix Example C-13). Often the position of a patient cannot be fully aligned with due to the institutional restraints of the GP, particularly in complex cases. The global requirement of balancing this dual stance can be seen to work at a local interactional level in these examples.

4.3.2.5   Checking agreement and understanding have occurred

Much as with establishing understanding in 4.3.2.2, a common technique was to explicitly check how much the patient understood of the consultation. This is done again by 'metacommunicating', or talk that gives information about the talk and here it is used with the role-playing patient (see Linguistic terms 6 Metacommunication). For example in '*MH 2*' (Appendix Example C-12), the candidate comments on the preceding explanation '`now i've gone through a hell of a lot there...have i made sense`'. Here, the patient responds well to the question, showing a stance affiliation with the candidate through his laugh, as well as following up with confirmation that the explanation has been understood. This checking, although good practice and expected as part of the exam model, can go wrong or be perceived as formulaic or patronising. Successful candidates will often mitigate these threats, customising questions with humour or more open questions. By contrast, pedagogic type requests such as '`can you just repeat for me what what are you planning today`' in the '*SH 1*' case (Misalignment Section 4.4.2.3 and Example 4-13) are likely to lead to misalignment.

4.3.3   Summary – alignment

- The term 'alignment' is more accurate and assessable than 'empathy' or 'rapport' in developing a shared language that is more concerned with evidenced behaviour than assumptions on people's inner feelings. 'Alignment' is useful in judging how the case is managed interactionally (interactional alignment) as well as assessing strategies of caring and agreement (affiliative alignment).

- Small differences at a local interactional level can have large consequences for how the consultation plays out and the performance of successful alignment.

- Alignment is a joint production between candidate and RP and the latter's' responses are a powerful channel through which a candidate comes to be judged.

- Alignment is happening throughout the consultation and often small and subtle cues in RPs' talk need to be picked up through accurate listening and inferencing.

- Other factors which affect the degree of alignment relate to the particular features of the exam i.e. the relative power of the RP, the patient-centred conventions of the exam and the need to rapidly imagine aspects of the case which have not been given in the notes.

- Successful candidates use conversationalising strategies more – small markers that make the consultation more informal and more sensitive to the listeners' stance. These include: making language more indirect, (e.g. using 'softeners' such as 'could', 'might') and more light-hearted.

- Affiliative alignment strategies function to show the candidate is patient-centred. These relate primarily to establishing patients' issues, concerns and expectations (ICE), to their psychosocial world and to check patient understanding. These 'empathy' tokens correlated with success but only where they were located appropriately in the turn by turn interaction and customised to avoid sounding formulaic (see 4.5 below).

4.4   Misalignment and misunderstanding

We touched on moments where alignment is not achieved in 4.3. Here, we look further at where communication goes wrong: communicative misalignment, missed opportunities for alignment, and misunderstandings – all of which might be called 'communicative trouble' (Schegloff 1987).

```
Example 4-10 – 'SH 1', IMG, Male, FAIL
 1 CAN: yes just for a few weeks
 2 CAN: and then then then
 3 CAN: ⌈it would get probably quite better⌉
 4 RPL: ⌊i i thought you said a for- month a few months⌋
 5 CAN: er yeah i mean at ten twel- mon- er weeks
 6 CAN: ⌈um i just⌉ broke it down to to weeks
 7 RPL: ⌊ok⌋
 8 CAN: ⌈but as i said that⌉
 9 RPL: ⌊sorry i'm⌋ confused there you said
10 RPL: ten twelve weeks and then you said two weeks so
11 CAN: no no i said twelve weeks sorry
12 CAN: i i'm sorry i'll repeat er (1.1)
13 CAN: ⌈the coil initially⌉ can make your er bleeding worse
14 RPL: ⌊⁇ok⁇ ⌋
```

'Misunderstandings occur whenever there is insufficient understanding for both parties to continue, where there is the illusion of understanding which is only revealed as such later on' (Roberts et al 2005; 468). They may arise from slips of the ear or from a more profound problem of understanding. Example 4-10 opposite will be addressed again throughout the chapter as demonstrative of the multiple factors at play in misunderstandings (MUs). Here though, we can see MUs flagged at lines 4 and then again at 9-10 as a clear misunderstanding of the preceding discourse, making such instances easily identifiable for analysis. More protracted negotiations

of understanding over many turns are also analysed, as well as instances where false starts, overlapping talk or the exam context contribute to confusing moments. Here we focus only on misunderstandings identifiable in the dialogue, much as the examiner must do, rather than any guessed 'latent misunderstandings' (Hinnenkamp 2003).

The opposite of 'alignment', misalignments are realised as uncomfortable and inappropriate moments or when one side expresses difficulty in interpreting the assumptions of the other. These include: 'social discrepancies where problems of sustaining social interaction lead to uncomfortable, disruptive or confusing moments and the notion of 'trouble' in interaction.' (Roberts et al 2005; 468). Misunderstandings can themselves cause misalignment: the troubled and socially discrepant moments which prevent the interaction from progressing smoothly. The 'missed' moments for alignment were, typically, exam-constructed moments where a role-player expressed something difficult or upsetting, which was then not picked up on by the candidate as requiring acknowledgement.

Altogether we coded 145 moments of misunderstanding, misalignment, or 'missed' 'empathy' cues, the results of which were given in section 3.5. Of these, 105 were misunderstandings, which were more prominent in the failing cases. Nevertheless, nearly all candidates had to deal with misunderstandings, so they are important for everyone to have a grasp of.

4.4.1   Who flags up the misunderstanding?

Most misunderstandings could be identified through the explicit expression of confusion by a speaker. It is interesting to add up which speakers generally do this:

Figure 8

Examples of misunderstandings

1. **Candidate flagged misunderstandings**
   ```
   CAN: just erm a b- it's it's a bit confusing just coming
   back in your family is there anybody else who is ********
   you mentioned your sister's children
   ```
   ('*JK*' Case, IMG, female, FAIL)

2. **Role-player flagged misunderstandings**
   ```
   RP:    sorry i'm just a bit still like wondering how i got
   got this  so ¬
   CAN:   mmm
   ```
   ('*JH*' case, UKG (White), male, FAIL)

3. **General trouble in interaction**
   Occasionally interactional sequences resulted in confusion, without either party seeming to express the need for repair. For example, the false starts from the candidate flummoxed by missing a physical examination card.

As can be seen in Figure 9, the majority (61) were highlighted by RPs accounting for over half the 105 misunderstandings. We know in 'real-life' GP consultations, misunderstandings often go unremarked on by patients, where highlighting lack of clarity by the GP is more face-threatening (Britten et al. 2000). In Roberts et al (2005) it is the doctor who nearly always flags them up, suggesting the most powerful in the interaction will take the initiative. In simulations, it is the more powerful RP who is performing this task.

Misunderstandings can also be cleared up gradually in real consultations, over the course of talking, rather than the overt, immediate repair required from RP-flagged CSA misunderstandings. It is of course part of the RP's responsibility to highlight any unsafe moments. While we saw in 3.5 that the behaviour of the role-player in 'sinking' or 'saving' does not correlate to case mark, it does seem significant that misunderstandings, which do show a pattern with success or failure, are largely flagged up by RPs. This gives RPs a large responsibility in what is highlighted as unclear and this is unsurprising since it is part of their role to indicate any lack of clarity.

Figure 9

Who highlights misunderstandings?

Candidate highlighted MU	27
Role-player highlighted MU	61
General trouble in interaction	17

4.4.2   Causes of misunderstanding

More tricky than noting who highlights the misunderstanding is working out why it has come about. Many misunderstandings are multi-factorial and a range of features work together to cause the confusion. Nevertheless, it is useful to break them down into categories.

4.4.2.1   Pronunciation, word-stress and intonation

Overt misunderstandings were not frequent in our data, perhaps inevitable in a sample of highly trained English speaking role-players interacting with candidates expert in English. There was no evidence that pronunciation caused MUs for either speaker. This contrasts with Roberts et al' study (2005) of real consultations, but in that study it was doctors' difficulties with patients' language that caused most misunderstandings. Occasional misunderstandings from a candidate's intonation could ensue. This candidate (Example 4-11) from the opening example places equal stress on the repeated words 'to to' at line 6, so that it sounds to the role-player as though he is saying 'broke it down to two weeks'.

```
Example 4-11 – 'SH 1', IMG, Male, FAIL
4 RPL: |i i thought you said a for- month a few months|
5 CAN: er yeah i mean at ten twel- mon- er weeks
6 CAN: [um i just] broke it down to to weeks
7 RPL: |ok|
8 CAN: [but as i said that]
```

This mishearing compounds an ongoing misunderstanding between the two speakers, and the case goes badly. Intonation and other prosodic features also occasionally lead to misalignment

in other areas. Most notably: in showing how explanations cohere together (see Explanations below); in functioning to alert the listener to shifts in the speaker's purpose (e.g. from explanation to data gathering, see below and differences in intonation may mask these shifts); and when candidates are perceived as sounding formulaic (see below 4.5.3)

4.4.2.2 Mishearing and misremembering

Some minor misunderstandings ensue simply from a speaker misquoting something the other has already said. It is often difficult to tell whether this is due to mishearing or simply misremembering the other's words. In Example 4-12, the candidate must apologise at line 55 after mishearing how long the patient's headaches have been going on for.

```
Example 4-12 – 'MCA', UKG (White), Male, PASS
51        mm okay just happening over the past two weeks, never had
52        anything like [this]
53 RPL:                [no no] not two weeks
54        a few weeks it's ⁇
55 CAN: a few weeks oh sorry I misheard you [there] yeah
```

Mishearings can often quickly be resolved. Nevertheless, they again evidence power differentials at work in the micro-level interactions of the CSA: since it is the role-player who usually points out the misunderstanding (see 4.4.1), it is therefore usually the doctor who must communicate the apology. In day-to-day GP encounters it is much more usual for the patient to be apologising for a mishearing, even where this mishearing is the doctor's (e.g. Appendix Example C-16).

When not quickly repaired, mishearings or misrememberings can lead to major misunderstandings. In a failing case for a lower scoring UKG (extract Appendix Example C-17), she forgets a piece of information that comes early on about a patient's family member not being a blood relative. The discussion of this family member's cancer leads the GP to offer a referral. A protracted renegotiation of understanding must happen five minutes in, when this forgotten information becomes apparent. Protracted misunderstandings always matched lower marks and it seems vital to repair missed information as soon as possible. However, corrective work must still be placed carefully, as 4.4.4 explores. Of course, it is not always possible to immediately realise a mistake has been made, and consideration could be given to allow more space for repair in the CSA, as suggested in 3.2.3.

4.4.2.3 Misunderstanding and misalignment from exam-modelling

The frequent phrases and strategies identified as part of the CSA exam model (see 3.4 and 4.3) can be used in a way that will backfire or cause confusion. In the comparative analysis of Example 4-6 from 'JA 1' and 'JA 3' (see 4.3.2.2), a difficult instance of establishing the patient's knowledge of cystic fibrosis and genetics resulted in minimal answers and a slightly discrepant, `i wouldn't know where to start'` (line 253) from the patient. In another 'JA' case, the same RP was more forthcoming on prior knowledge (Example 4-7). This strategy was clearly successful for many of the candidates discussed in section 4.3.2.2. For the IMG in

Example 4-6, perhaps it is down to the lack of any softening/politeness markers in the first part of the question '`do you know much about chromosomes`', although many other candidates in fact get away with much more unhedged questions, such as *'MS 2'* '`what do you know about glucosamine?`' See Example 4-2. It may also be that this candidate's exam-modelled ICE questions occurs alongside many other alignment difficulties and misunderstandings during her case, meaning the RP is less amenable to the ICE question here. The line between which exam-style patient-centred questions work well or fail is therefore quite fine, and they need to be used with some care, as section 4.5 on formulaic language explores further.

Open-ended ICE enquiries also open up the opportunity for the patient to express misunderstanding. While such avenues might be good practice for the patient-centred consultation, in the exam context we are presented with the obvious strategic problem that any expression of confusion highlights problems. The exam modelled 'checking' question initiates a misunderstanding sequence in '*SH 1*', a case already noted to contain a protracted misunderstanding in Example 4-10, (see 4.4), and given more fully here in Example 4-13. At 8 minutes in, lines 422-5, the candidate asks;

Example 4-13 – '*SH 1*', IMG, Male, FAIL

```
422   CAN:     j⁇xx⁇ (1.3) just to
423   CAN:     make sure that i've made myself clear i mean
424   CAN:     could you quickly repeat
425   CAN:     what wh- wh- wh- what are you planning today
427   RPL:     you're going to give me (.) well (.)
428   RPL:     nothing really in [that you're] giving me some [information]
429   CAN:                        [uhmm]                      [yeah]
430   RPL:     to take away [about]
431   CAN:                  [yep]
432   RPL:     the pill which you said
433   RPL:     which sounded e- w- was like
434   RPL:     um the one i took in my [twenties]
435   CAN:                             [could be yes yes]
436   RPL:     but wasn't great [if i was still smoking]
437   CAN:                      [no yeah]
438   RPL:     an um (.) then
439   RPL:     there's a coil which works
440   RPL:     as both a contraceptive
441   RPL:     [if i want to reduce my periods but it would make] them≈
442   CAN:     [and ⁇um⁇ yes]
443   RPL:     heavier to start with
444   CAN:     yes just for a few weeks
445   CAN:     and then then then
446   CAN:     [it would get probably quite better]
447   RPL:     [i i thought you said a for- month a few months]
448   CAN:     er yeah i mean at ten twel- mon- er weeks
450   CAN:     [um i just] broke it down to to weeks
```

Micro-level analysis of the 40 cases

This question is more pedagogic in nature and not mitigated in the same way as the ICE requests described in 4.3.2.2, and 'made myself clear' while intended to be face-saving for the RP patient, could have resonances more associated with a reprimand as in 'I hope I have made myself perfectly clear'. The question is met critically by the actor; '`you're going to give me well nothing really`' (lines 427-8). This is not the only factor which feeds into the protracted misunderstanding: we saw in 4.4.2.1 there was an intonation/mishearing issue, as well as further causes explored in 4.4.3. Nevertheless, it is this rehearsed ICE question (perhaps from formulaic training and preparation) that initiates the most overtly problematic sequence of the case. This shows again some of the tensions presented by an exam setting, where modelled patient-centred questioning does not always present the most successful exam strategy.

4.4.2.4   A note on cultural misunderstandings

Much as in Roberts et al.'s (2005) study of primary care consultations in Lambeth, misunderstandings stemming from differences in culturally specific health beliefs did not occur often. Only one instance in the data could be said to demonstrate a mismatch in cultural knowledge – in which a candidate explaining genetic inheritance asks a white RP about first cousin marriages as if this was a routine question to ask any patient.

There are some sources of misalignment that could tenuously be attributed to differences in cultural knowledge, such as an IMG candidate's lack of discussion about the relative merits of school dinners versus packed lunches for the teenage *'KM 1'* (see Appendix Example C-18) or the ability to rapidly imagine contextual details that are persuasive to an noncompliant patient (see discussion 4.3.1.2). However, such moments do not lead to overt misunderstanding. The absence of overt cultural misunderstandings from the data (experienced either by RPs or candidates) is perhaps not surprising given that the great majority of cases and RPs do not represesent patients from very different minority cultural backgrounds.

4.4.3   Misalignment

As we saw in 4.4.1, there can be generally troublesome or socially discrepant moments which prevent the interaction from progressing smoothly that no one flags up as requiring repair, but that might influence the sense of what was generally called 'clunkiness' in feedback sessions.

4.4.3.1   Disfluencies and trouble

There are often numerous false starts and hesitations in candidates' talk, and many coincide with cases that are problematic or have more overt misunderstandings, as part of the general finding that misalignments were more common in these cases; for example with a UKG (Example 4-14). False starts and repetitions are a feature of any conversation and all candidates experience them at some point. Stammering can be registered as a disability and the College will make reasonable adjustments to any candidates registered as such. Other forms of disfluency do not fit this category.

The nervousness and artificiality of some aspects of the simulation could certainly be argued to contribute to disfluency. For example, a highly marked female UK graduate (see Appendix Example C-19) gets up to do a physical exam with her back to the examiner, at first not hearing the examiner's intervention, resulting in hesitation and a few false starts. But what seems significant here in a context in which talk is assessed, is how much more they might mean to the observer, particularly since they occur more often in those cases where the candidate experiences other misalignment problems. They disturb the rhythmic co-ordination of talk, which may frustrate the RP and even lead to other problems within the interaction. These very small level interactional features, some of which are a product of the exam, can, when they occur repeatedly, build up a more general impression of a poor candidate (see also 3.5, 4.3.1.1., 4.3.1.3, 4.3.2.2, 4.3.2.3 and 4.4.2.1). These small moments of trouble are amplified in an exam setting and our data show that candidates loose marks because of them and for somewhat weaker candidates may be the difference between pass and fail, although it is not possible to establish this definitively.

```
Example 4-14 – 'SH 2', UKG, Male, BORDERLINE FAIL
54   CAN:  over sort of how long of a period
55   CAN:  they've started t- t- to (.)to to (.) to (.) to get worse
56   CAN:  do [you think]
57   RPL:     [ um]
58   RPL:  i would say really its been the last year now
59   RPL:  that i've been getting quite fed up with them
60   PPP:  (0.7)
61   CAN:  okay (0.4)
62   CAN:  um (1.0) and (1.3)
63   CAN:  you mentioned you always hav- you've always had heavy [period]s
64   RPL:                                                          [ yeah]
65   CAN:  when er when did you start your period [if i if i may ask]
66   RPL:                                         [i was age thirteen]
67   PPP:  (0.7)
68   CAN:  thirteen okay
69   CAN:  um (0.5) and
70   PPP:  (1.0)
```

4.4.3.2   Causing patient alarm/anxiety

Other moments that might indicate a poorer relationship between candidate and role-player are those where alarm is voiced. In Example 4-15 from the *SH 1* case, the candidate has requested they do a physical examination and the patient asks for further clarification. Unlike the majority of the physical examinations in the CSA, this candidate has not prefaced his request with a description of what he will be examining. In the simulated CSA such prefacing work is vital, indicating to the examiner as well as the patient what kind of examination is intended, which, given that it is often not carried out 'for real', may be the only opportunity to demonstrate competence. This important 'metacommunication' work was discussed in Linguistic terms 6, as crucial to CSA talk. The role-player's request for more information seems to cause some nervousness, again with many false starts (line 180 'w- we we we look for any

```
Example 4-15 – 'SH 1', IMG, Male, FAIL
173   CAN: ≈er and er i mean er
174        a quick er b- abdominal examination
175   tim: 3.30sec
176   CAN: just to make sure that i mean you're not missing any any
              anything
177        |serious there|
178   RPL: |what are you| looking for
179   CAN: er usually er i mean
180         w- we we we look for any any
181         abnormal growth let's say i mean
182        [i might like to ⁇ ]
183   RPL: |that sounds quite worrying|
184   CAN: no it's not at all no
185        er b- i mean i i haven't phrased it right i would say
186        er it's just to make sure that i mean er your kidneys are fine
187        your womb is ok i [can't feel it|
188   RPL:                   |right|
189   CAN: an- is is
190   RPL: and kidneys might cause this
191   CAN: er kidneys no the- they wouldn't
192        [it it'll be just a part of| a general exam[ination]
193   RPL:  |yes ⁇|
194                                             |oh ok|
```

any). His brief explanation that it is to 'look for any any abnormal growth let's say', is not mitigated with any of the usual caveats in trying to minimise alarm, talking about likelihood or ruling out the worst case scenario. This is perhaps understandably met with the patient's metacommunicated concern that it 'sounds' alarming. The candidate metacommunicates back that no 'er b- i mean i i haven't phrased it right i would say' (185). He then tries to explain it further but makes a further mistake in suggesting he is checking 'kidneys'. One wonders here if he is simply listing different organs in the abdomen (kidneys, womb...) in an attempt to make it all sound more routine and less alarming, a more 'general examination'. Nevertheless starting with the mistaken 'kidney' examination clearly causes further problems, and the roleplayer is easily able to cut in through one of the candidate's false starts (189 an- is is) to flag up this mistaken clinical direction: 190 'and kidneys might cause this'. His failure to initially explain the reasons for needing a physical examination leads to this disrupted structure, poor explanation and the articulated patient alarm. This in itself is an example of how interactional and multifactorial misalignments can be, but along with the other resultant misunderstandings of the case show how accumulatively they can function. This misaligned interaction clearly works toward the larger breakdown between role-player and candidate already noted in 4.4.2. Along with other difficulties, this example illustrates the importance of developing good, clear explanations both for real life and the CSA.

Candidates that implied even more minor moments of alarm with patients were those from failing cases, as given in some further examples in Appendix Example C-20. These are the negative face of the alignment strategies in 4.3 above that demonstrated reassurance,

downplayed bad news and maintained a more conversational, sometimes light-hearted approach.

4.4.4 Repairs

4.4.4.1 Repairs – Missing crucial questions and information

Often when candidates take the wrong course in a consultation, offering up an unsuitable diagnosis or plan, they can correct themselves by asking importantly missed questions later. The passing UKG candidate in the '*RT*' case (Appendix Example C-21), was already noted in section 3.2.3 revisiting his diagnosis and explanation phase. He places a forgotten question after the end of his first explanation phase; 'the other thing i i spec- i forgot to ask you earlier on was do you ever get a sensation of something coming down down below um', successfully transitioning into further data-gathering. This in turn requires a revised diagnosis and new treatment plan. He passes the case, although still loses marks for organisation.

Not all candidates are successful in a shift in diagnosis part way through. Where missed questions are placed within the local interaction seems particularly important. The unfortunate '*SH 1*' candidate, who places his forgotten question in the middle of an explanation phase ('CAN: er one thing which i forgot to ask i mean is it painful when when...' see Appendix Example C-22), without differentiating it through his intonation, is much less successful in repairing the mistake.

Explanation phases are a particularly important part of the consultation. They require a good deal of talk from the candidate and, as we shall see in 4.6, the careful structuring of information. This self-interruption by the candidate during an explanation therefore contributes to this 'troubled' interactional phase. On occasions where the role-player plays a heavy part in putting the candidate back on the right path, the candidate also still seems to do badly, despite ostensibly repairing the mistake.

4.4.4.2 Self-corrections

Given that actors flag up misunderstandings more often in this setting than in a regular GP surgery (4.4.1), it is interesting to note that many successful candidates slipped a self-correction into their turn before the role-player has had a chance to respond. These were tagged as GP initiated repairs, although it might be difficult to describe these short instances as full misunderstandings. Example 4-16 from a high scoring candidate in the '*TJ*' case shows a successful self-correction, mid-turn, with a further example in Appendix Example C-24. As noted in the introduction at 4.3, she quickly corrects the potentially discrepant word 'disguise', which may threaten the positive view the patient holds of having a secret vasectomy, in order to demonstrate some affiliation to the patient's stance. Self-initiated repair can be a useful tool then. What seems again important about these is the very local interactional level at which they are achieved, before anyone might even notice a slip-up. Repairs carried out after a misunderstanding has been flagged, such as the previous section on missed questions, become, in our data, much more likely to impact negatively on a candidate's marks.

```
Example 4-16 – 'TJ', UKG (White), Female, PASS
315 CAN:  and so so it is er it is
316 CAN:  quite a lot |to consider before| you go for an operation
317 RPL:               |mmm|
318 CAN:  like this i mean in some respects
319 CAN:  it's very easy it's quick
320 CAN:  but in others it can leave you with
321 CAN:  even |long-term consequences|
322 RPL:       |yeah yeah| yeah
323 CAN:  and i just wonder if if you're not going to tell your (.) partner
324 CAN:  how you might (.)
325 CAN:  disguise some of those complications should they arise
326 RPL:  |umm|
327 CAN:  |i'm sorry| disguise is maybe the wrong word to use
328 CAN:  but (.) you know what i mean
329 RPL:  yes um i suppose i'd
330 RPL:  i'd i'd face those things as and when but um
331 RPL:  (.) i'm
332 PPP:  (1.3)
333 RPL:  perfectly willing to do that
```

4.4.5 Summary – Misunderstandings, misalignment and repairs

- Virtually all candidates experienced moments of misunderstanding and/or misalignment, but poorer performer candidates had slightly more.

- RPs flag up misunderstandings in the CSA more than patients tend to in real life GP encounters and GPs tend to be on the back foot apologising, demonstrating the different interactional setting.

- Misunderstandings were caused by mishearing and forgetting, missing clinical information and exam-modelled sequences that 'back-fired' in terms of RPs' responses. Other misalignment could be caused by disfluencies in speech and causing patient alarm. These could compound each other in multifactorial misunderstandings. Disfluencies occur in all talk but are more likely in the exam context, raising the question of whether a position of communicative tolerance could be adopted.

- Much of the misalignment results from a lack of metacommunication and mitigating language. Special attention should be paid in CSA preparation to explanations of: why a physical exam is being undertaken, what is routine and what is a possible cause of alarm.

- Attempting to repair trouble was difficult. Where the trouble was repaired swiftly and located appropriately, it could be rapidly defused. In all other cases, it was difficult to recover from mistakes without losing at least some marks, as was discussed for late data gathering in 3.2.3. One recommendation might be to consider further opportunities for candidates to repair mistakes – interactional misunderstandings will be a common occurrence in everyday GP encounters and professionally we all need to know how to

repair them. Again this is a suggestion for more communicative tolerance. In this assessed setting, it seems possible they will stand out more than usual, especially where overtly 'metacommunicated', with the emphasis on the mistake, ('sorry one thing which I forgot to ask'). In interaction there will always be disfluencies and some level of less than complete understanding. The important element to focus on is the ability to negotiate understanding to prevent and repair misunderstanding and misalignment (Gumperz 1982) and this negotiation rather than the original trouble is what should be focussed on.

4.5 Formulaic language and (mis)alignment

In the last sections, we looked at sequences of speech with important socio-interactional functions in the CSA; questions that elicit 'ideas, concerns and expectations', ways of balancing 'empathy' or approval for the patient's stance with institutional constraints and explicitly signalling these functions through 'metacommunication'. Many of these were seen to fit with a broader communication skills model of the consultation, such as Neighbour (1987) as well as the Calgary-Cambridge model (Kurtz et al. 1998). However, these sequences were often commented upon by examiner feedback as sounding 'formulaic' and 'learned' (see Chapter 5 and quotes Appendix Example C-25). A crucial question we raised, therefore, was why some sequences should stand out as jarring and inappropriate.

Successful candidates were in fact found to be using more 'exam-modelled' strategies than poorer performing ones (see 3.5). The quantitative analysis showed close similarities in the phrases used (3.4), as well as many of the alignment strategies (4.3.3). It is hardly surprising that all candidates do this, given the direction to perform these 'phrases' from the medical communication literature ('Work out exact phrases which demonstrate empathy in specific situations.' from Kurtz et al. 1998: 134). Since, on the surface, it is hard to find differences, there must instead be considerable interpretive weight to the way these phrases *sound* in *context*.

So why were some candidates assessed as *sounding* more formulaic? Micro and corpus analysis from our data show three main reasons, explored here: the design and extent of 'empathy' phrases, their location in the talk and how they sounded or were supported (or not) by bodily movement.

4.5.1 The design and extent of 'empathy' phrases

Let's take the word 'feel' as an example of how there come to be 'preferred ways' of phrasing things. We saw in 3.4 that 'feel' occurred in the top 10 keywords and the phrase 'do you feel' was the top 3-word cluster when compared with general spoken business English. Appendix Table C-1 shows the phrases that 'feel' tends to occur within, according to the most frequent words found either side. Most of these were utterances that examiners themselves identified as CSA stock phrases in feedback sessions; 'how do you feel about....', '...make you feel', 'do you feel....'. There are then clearly recognised 'formulaic phrases' in the CSA. But since they occur

in nearly all cases, they become a difficult feature for differentiating good and poor candidates.

One difference commented on in examiner feedback relates to slight modifications of the most stereotypical phrases: an instance where 'how do you feel about that' which was thought more acceptable than the stock, 'how does that make you feel?'. We saw throughout 4.3 other strategies for customising these formulaic exam phrases, such as casualising the interaction by inserting in 'conversational' softeners (e.g. kind of/sort of; 'do you mind if I ask you a few sort of personal questions', '*JH*' case). Being able to 're-formulate' stock phrases is a key feature of conversational discourse, and requires some implicit confidence with the language genre to be able to do so (Schmitt and Carter 2004, Carter 2004). Other ways of customising exam strategies is to preface them using other conversational formulas, such as 'do you know what...' in '*MS 2*' (see 4.5.3, Example 4-19) below, or minor swearing 'now i've gone through a hell of a lot there' from '*MH 2*' in 'checking understanding' (see 4.3.2.5). These strategies make such utterances appear to stand out less and be less exam-modelled and formulaic, in contrast to their use by some IMGs, discussed in 4.5. When formulaic utterances are made to stand out in everyday, conversational English, this is usually for creative effects such as irony or to make jokes (Carter 2004), perhaps accounting for why they seem insincere when they stand out in the CSA.

Those who know the 'rules of the game' for a hybrid of casual chat/ more formal and institutional talk of British GP consultations, are going to have greater facility with re-formulating the stock phrases of the CSA.

The other difference was the way 'feel' phrases were used to relate to different topics: emotions, thoughts or physical symptoms. A sample of these can be seen in Appendix Table C-2. Successful IMG candidates often don't just use 'feel' phrases in relation to emotional wellbeing but also when eliciting physical symptoms (e.g. 'do you feel any palpitations'). IMG candidates who use it only in relation to emotions appear to be those doing less well overall. While a tricky inference to make, perhaps the use of 'how do you feel' in relation to emotions sounds less 'marked' (Linguistic terms 8) when it is also used in other contexts. A facility with crafting how CSA phrases are used seems important.

4.5.2   Location in the talk

We've already looked at the importance of the micro, turn-by-turn sequencing of talk, and particularly how rapid topic shifting away from upsetting topics could cause misaligned sequences (4.3.2.3). 'Empathy' questions, where they are interjected between a series of clinical questions and when not elaborated on, can also be seen as brief or tokenistic.

A clear example is the '*SH 1*' candidate (who receives a 0 for the case and examiner feedback comment number 11: 'Does not appear to develop rapport or show sensitivity for the patient's feelings'). In Example 4-17, line 147 '`Yeah no I do understand`' is an 'empathy' phrase often employed in the CSA, but here is immediately followed by further biomedical questions, so was rated as 'clunky'. The picture is not completely clear cut though. Brief

```
Example 4-17 – 'SH 1', IMG, Male, FAIL
133  CAN:      what kind of job d'you do
134  RPL:      i'm a primary school teacher
135  CAN:      uhm and is it affecting you work
136  RPL:      hh- well
137  RPL:      i mean i'm still manage to do everything
138  RPL:      but i do worry about flooding
139  RPL:      ⌈you know in front of the class⌋
140  CAN:      ⌊yep yeah⌋
141  RPL:      ⌊you know⌋ it hasn't happened ⌈but it is a concern you know⌋
142  CAN:      ⌊yeah⌋                        ⌊mhmm mhmm⌋
143  RPL:      and you can't just run out of a class ⌈mid way⌋
144  CAN:                                            ⌊no⌋
145  CAN:      ⌈no i know⌋
146  RPL:      ⌊is to change a⌋ pad or something
147  CAN:      yeah no i do understand
148  CAN:      and and do you smoke
149  RPL:      yeah not very much
150  CAN:      how many a day
151  RPL:      er about ten (0.7)
152  CAN:      ten a day ⌈and how long⌋ have you been smoking for
```

expressions of 'empathy' or understanding can be acceptable, such as the '*MH 2*' candidate in 4.3.2.3 who quickly moves away from the patient's upsetting description of his father's illness to saying '`sure I understand··· ·yeah i mean the the good news about haemochromatosis are is···`'. He smoothly transitions from expressing understanding into his biomedical explanation by metacommunicating the positive information he is about to give this patient (see also Appendix Example C-9 and 4.6 on Explanations). This kind of metacommunication helps these quick expressions of 'empathy' before topic changes to be carefully navigated, and to better acknowledge the patient's previous description of difficulty. Again small differences at a very localised interactional level are important to the more general impressions built up of sincerity and alignment.

4.5.3   How 'empathy' phrases sound

Successful candidates have, as indicated, several ways of washing out the formulaic in 'empathy' phrases; customising the wording a little, using formulaic words such as 'feel' in other contexts so they lose their recipe quality and carefully locating them. Another important aspect depends upon how they sound, or their 'prosody' (see Linguistic terms 7).

Two candidates are contrasted below in the way they deliver expressions of 'understanding'. The first does well overall, passing the CSA and even a borderline pass in this case. However it is notable that he does particularly badly in the Interpersonal Skills domain, both as his lowest mark overall (24/39) and in this case, where he scores 1 and receives the feedback: '11. Does not appear to develop rapport or show sensitivity for the patient's feelings' and '13. Does not make adequate use of verbal and non-verbal cues, Poor active listening skills'. This is a complex case where the mother is concerned her young son is being abused by her ex-husband's brother, a full extract for which can be seen in Appendix Example C-26. The candidate performs many

of the explicit affiliative alignment strategies discussed in 4.3.2, expressing care '`sorry to hear about that`' (line 52), as well as his understanding for the RP's situation (lines 100-102, Example 4-18):

Example 4-18 – 'JG', IMG, Male, BORDERLINE PASS

```
100   CAN:        right i can see why you are concerned
101   CAN:        and er it's not something to
102   CAN:        to be taken lightly and i can see where you're coming from
103   RPL:        mmm
```

These are typical CSA strategies, very similar to the function of an interactional sequence in Example 4-19 (full extract Appendix Example C-27), from a candidate who does extremely well (and see the analysis below).

Example 4-19 – 'MS 2', UKG (White), Female, PASS

```
481 RPL: it does seem very unreasonable an- ≈
482 CAN: +≈ yeah
483 RPL: whereas if you
484 RPL: almost if you cross the road they say oh that's fine
485 CAN: yeah
486 CAN: no i do do you know what i really do understand that
487 CAN: particularly if you've got friends who…
```

In analysing the intonation patterns candidates use to show understanding, we can make some claims about how things sound. While difficult to represent visually, the graph in Figure 10

shows the volume (in the green line with jagged, peaked contours) and the pitch (blue line) as the IMG candidate says;

```
'Right I can see why you are concerned and er it's not som ething
to be taken lightly and I can see where you're coming from'
```
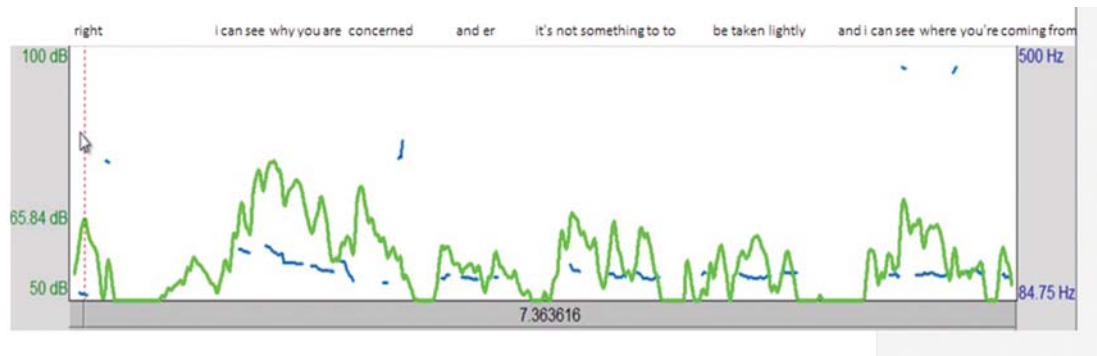
Figure 10 – Intonation and Volume in Example 4-18

Many of the IMGs, including this candidate, are from a South Asian background and use a system of English which draws on aspects of North Indian languages such as Hindi. Word order is more flexible in languages like Hindi and a lot of the work that is done in English with stress and intonation can be done by juxtaposing words and phrases. So information tends to be conveyed in small packages juxtaposed to each other as we can see here; `right / i can see why you are concerned / and er / it's not something to to / be taken lightly / and i can see where you're coming from`. Each unit uses the same melody, with the volume rising and then rapidly falling so it sounds like a list and perhaps rather formulaic. And while the volume is raised at the beginning of each information package, the pitch remains quite low and flat. In local English this could sound uncaring but in Hindi and in Indian English low pitch is conventionally a marker of respect or conveying bad news (Gumperz 1982).

In local UK English, we saw that information units are produced in smooth envelope contours (see graph in 'Linguistic terms 7' and more fully Appendix Example C-28). The emphasis at the opening of the utterance, which is a little higher and louder, forms an 'affective contour'. In the UKG's turn at line 486, she has a false start (`no i do`) before she says (`do you know what i really do understand that`), with a conversationalising discourse marker (`do you know what`). This whole line, inclusive of the false start and the discourse marker, is enclosed in this one smooth envelope contour without a pause, with a slight volume emphasis on 'do' and 'really', which rounds off at the end with a drop in pitch. So as well as customising the 'I understand' phrase, she gives it expressiveness with the higher tone. By contrast, much emphasis in Hindi and Indian English is done through little words called emphatic particles. Emphasis will not be done so much through intonation and utterances can sound flat.

Intonation and emphasis are forms of metacommunication (see Linguistic terms 6), helping to signal to listeners how you want your talk to be received, especially important when there is an

overhearer, such as the examiner in this case. All this interpreting of meaning and attitude goes on below the level of consciousness and is produced automatically, as a result of learning to interact with others who use features like us. This is then something that is not being explicitly assessed, but nevertheless may impact on how utterances are interpreted. In an exam where 'CSA type' exam modelled expressions dominate much of the talk, their prosodic delivery can be key to whether they stand out as 'formulaic'.

4.5.4   Gesture and facial expression

Comments were often made about candidates' gesture, gaze and facial expression in examiner feedback (Chapter 5). Contrasting two cases helps illuminate why some candidates might appear more sincere. This pair of sequences is taken from the '*MS 1*' and '*MS 2*' cases, at a point where, the RP describes how osteoarthritis is impacting on his life (full extracts in Appendix Example C-29). In Example 4-20 and Example 4-21 the RP talks about his dancing, which both candidates acknowledge with a non-verbal cue (for anonymity, the face of the failing candidate is not shown). There are differences though. The second candidate Example 4-21, rather than producing a head nod, smiles. Following this, the RP continues by describing the difficulties he now encounters at the dances, which elicit further non-verbal head nods, accompanied by a frown. Of course, no candidate will fail on the basis of individual gestures in the interaction, and indeed the examiner does not have time to consciously mark every smile. But built up over the course of a consultation, especially

```
Example 4-20 – 'MS 1', IMG, Male, FAIL
107 RPL: my wife and i (.) quite enjoy dancing↘ (.)
108 %ges: CAN nods
```

```
Example 4-21 – 'MS 2', UKG, Female, PASS
140 RPL: er I (0.4) my wife and I go dancing↗ um
141 %ges: CAN smiles (00:01)
```

from this early data-gathering stage, they do make a difference to impressions of affiliation. Gestural communication comes with some difficulties, and such involvement strategies can also be interpreted as 'over the top' and 'acted out', as discussed in Chapter 6.

4.5.5   Summary – Formulaic language

- Although most candidates use the same 'empathy' tokens, unsuccessful candidates are routinely judged as being formulaic and lose marks on interpersonal skills.

- Successful candidates use strategies to bleach out the formulaic and sound sincere: customising the wording a little, using formulaic words such as 'feel' in other contexts so they lose their recipe quality, not locating them briefly in biomedical sequences, using expressive non-verbal features.

- It is challenging for candidates whose English is influenced by other languages (and

this is the case for speakers of Indian English, among others) not to sound formulaic since intonation and prosodic features may be used differently to express affect and sound more 'marked' than UK graduates (see Linguistic terms 8).

4.6   Explanations

So far we have looked at isolated micro sequences of interaction across entire cases. In this section we look at a larger, self-contained phase of the consultation, 'explanations', to address how micro-interactional features work together simultaneously. We established in 3.3 that explanation phases represent particularly effortful 'talking work' for the candidate, during which time they hold an average of 81% of the conversational floor, providing a carefully structured piece of talk. While a disrupted turn-taking pattern accounted for a few poorer explanations, the majority looked identical on the surface. It falls to a closer analysis of what is said to understand where things go wrong. The high amount of talk required from the candidate would seem to work against the phenomenon we were describing in 4.3.2.3 above on the joint production of interaction and patient-centredness. But in fact many features in the successful consultations continue the sense of dialogue in this quite monologic phase, maintaining the patient-centred strategies that seem so important to success. Often a consultation would contain more than one explanation phase, meaning that in total we annotated and extracted 81 phases for analysis from the 40 cases.

4.6.1   Types of explanation

Just as we identified types of cases in 2.5, so too do explanations fall into different categories that roughly reflected these distinctions, the numbers for which are given in Figure 11.

For space, our analysis is here compressed into a single outline of common characteristics of explanations. The detailed analysis of a 'Demanding clinical explanation', highlights many of the communicative features that serve well for all types of explanation, and can be compared with poorer performing candidates to explicate communicative difficulties contributing to failure.

4.6.2   Common characteristics of explanations

The explanation can be analysed in terms of:

i)   Relationships between ideas – This concerns the rhetorical structure of the explanation to give information, such as using metaphors or contrasting ideas to help make concepts clearer, linking information together through repetitions and identifiers. Four discourse strategies for linking ideas together are given in 4.6.3 below.

ii) Relationships between people – These are features which make the explanation patient-centred (such as guiding a patient through the argument more explicitly, taking account of anxiety) and which show the orientation of the speaker. They include features such as;

Metacommunicating – signposting where the explanation is going and indicating stance (e.g. 'I don't think we need to worry about this'); Relating information back to patient concerns and ideas (e.g. 'you said that was painful' in responding to patient's assumptions), gliding between the clinical and the personal; managing anxiety and uncertainty. Three strategies are given in 4.6.4 below.

Figure 11 – TYPES OF EXPLANATION	
Routine clinical explanations This comprised the majority of cases. They described conditions ranging from frozen shoulders to more serious conditions such as Parkinson's. Nearly all tended to mark the culmination of a period of data gathering and/or physical examinations or test results.	**45**
Demanding clinical explanations Particularly indicative of explanations e.g. in genetics cases, where an intricately linked description of inheritance risk and potential conditions had to be made, often with several interlinked explanation phases. The cases did not necessarily involve the kind of agonistic or emotional decision making of the typically 'complex cases', but rather were difficult through the sheer amount of complex, logically linked information required, as well as taking into account patient anxiety.	**20**
Stance saturated These involved the doctor's institutional position as a key part of a joint decision making process, balanced carefully with the stance of the patient; e.g. the doctor in '*MS 2*' explaining why the patient cannot have Glucosamine on prescription, with many of the balancing strategies outlined in 4.3.2.3.	**10**
Social/emotional decision making Similar dilemmas to 'stance saturated', but tended not to include the stance of the doctor to the same degree since they involved a more personal decision by the patient; '*AP*' involves describing termination options to a young woman who is unexpectedly pregnant.	**6**
TOTAL No. of EXPLANATIONS	**81**

Successful candidates manage both aspects well, conveying information clearly and aligning their explanation to the listeners (RP and examiner). Here we analyse a minute-long explanation from a successful British male candidate ('*MH 2*' case, Clear Pass, 110 overall, see Appendix Example C-30 for full extract). While this is a more demanding, explanation heavy case than is common in the CSA, it demonstrates many features of managing relationships between ideas and people that are helpful even in more simple explanations. In this initial

explanation phase, he describes the specific characteristics and symptoms of haemochromatosis.

During this minute-long explanation we can see the RP only makes very short responses, always overlapping the talk of the candidate ('mmm', 'right', 'yes') rather than taking over the floor. The RP's role here is a supportive one while the candidate talks, but continues to signal a degree of involvement. This then is the kind of interactional alignment that was discussed in 4.3, where both parties implicitly agree that this is a particular phase, an 'explanation', during which the candidate should be granted the majority of the floor time. This is similar to research on narratives and the speaker's right to the floor (Sacks 1992). These explanation phases show other resemblances to narrative structures in conversation, as discussed further.

4.6.3 Relationships between ideas

4.6.3.1 Identifiers and mini-explanations

Signalling logical connections is essential to clear explanations. In the early stages of the haemochromatosis explanation, definitions identifying the disease are given. These rely on linking the named disease to mini

Example 4-22 – Identifiers and mini-explanations
```
265 CAN: you know haemoˈchromatosisˈ that y-
266 ACT:                    ⌊mmm⌋
267 CAN: it is (.)
268 CAN: what we call a um i mean it's a genetic condition
269 ACT: mmm
270 CAN: um which means it is inherited
271 ACT: right
```

clarifications through subordinate clauses: 'which means', to link the identified disease 'haemochromatosis' to its characteristic 'it is inherited'. These mini-explanations are given in the smooth envelop contour packages addressed in Linguistic terms 7: '`which means it is inherited`' all forms one unit, following on from a similar contour '`I mean it's a genetic condition`'. Even with the minor disfluencies that are typical in all talk, this candidate, through his words and prosody, links information together.

However, much as with the formulaic utterances discussed in 4.5, it is possible that some IMG candidates sound 'marked' (see Linguistic terms 8) when they use this structure for giving linked information. A male IMG candidate from the '*DT*' case (see full explanation Appendix Example C-31) receives a 3 for the case and a 1 in Interpersonal Skills, with the feedback '16. Does not use language and/or explanations that are relevant or understandable to the patient'. His mini-explanations are slightly different. He introduces the anatomical term at line 240 '`you know like that part of the knee where is the meniscus what we call`'.

He is using the same kinds of phrasing, but in a way slightly different enough to make it sound marked. Placing the medical discourse maker, '`what we call`', after 'meniscus' has the effect of producing three separate phrases which are not connected as a smooth prosodic contour, i.e. `that part of the knee/ where is the meniscus/ what we call` (instead of 'that part of the knee which we call the meniscus'). Again a little later in the explanation he uses the subordinate clause '`that you've damaged your`

`ligaments which are you know sort of sport injury`' (264-5), but the logical connection between damaged ligaments and sport injury is not made grammatically. Prosodic differences in another failing male IMG candidate (see extract Appendix Example C-32 from '*JL*') may also contribute to his poorer explanation and low grade. At lines 161-3 (`haemoglobin is twelve point one /which is slightly on the (.) lower end`') and 179-83 (`also your f-s-h / which is a one of the hormones that goes up when you have a um when you are (.) having the change`'), there are pauses in the middle of the delivery of these mini-explanations and a slight raise in pitch afterwards. The link between the statement and the subordinate or relative clause ('`which...`') is not easily processed and seems to disrupt the logical connection. So, despite using many of the strategies of successful candidates, small differences that only show up at this micro level can make it sound marked and more confusing.

4.6.3.2 Repetition

Logically connecting information can also be achieved through listing strategies. In the haemochromatosis case, the candidate lists evidence by describing symptoms, in Example 4-23. The repetition of the '`it can`' structure works strategically well for conveying a lot of related information in a short space, and can be used to signal a conclusion '`eventually can lead to kind of liver failure`'. If we look at the intonation patterns he

Example 4-23 – Rhythmic repetition for listing concepts

290	*CAN:	which are the things i was asking you so **it can** actually cause it
291	*PPP:	(0.5)
292	*CAN:	**it can** make you look very ↗tanned
293	*ACT:	mmm
294	*CAN:	**it can** affect your ↗sugar ↘levels **it can** affect your ability
295		ability to get an ↗erection **it can** affect
296	*PPP:	(0.5)
297	*CAN:	your your you know the way your ↗liver works
298	*CAN:	and then ↗[event]ua↘lly
299	*ACT:	[yeah]
300	*CAN:	↓**can lead** to kind of liver failure

uses when giving this list, we can see how this list is performed within the narrative-style structuring of the explanation. At the end of each item, his intonation goes up, signalling the new piece of information given – '`it can make you look very⤴tanned`'.

Much as was introduced in 4.5.3, this again, is a typically British English prosodic patterning – up and down, with each unit contained in a prosodic contour and each new information word emphasised as important. His pitch then trails downwards at the end, signalling the conclusion – '`eventually can lead to kind of liver failure`'. Not all repetition is used strategically or works well but it is particularly useful in this explanation for highlighting symptoms and causes.

Some IMG candidates exhibited different intonation patterns with the stress falling on different items for emphasis. For example, the IMG giving the '*JL*' explanation (Appendix Example C-32), uses a volume and pitch rise to emphasise a mistake. He is listing symptoms that could have affected the results of the thyroid test; '`unless you had any any problems when you had the blood tests (1.2) cold sore throat sometimes sometime not sore throat flu like symptoms which can pain around your neck an you can have thyroiditis`'(lines 306-13). However this list – 'cold', 'sore throat', 'flu like symptoms' – is given in a flat, level pitch and volume (see pitch contour Appendix Example C-33). The only rise comes when emphasising a mistake he feels he has made; '`NOT sore throat`'. Emphasising the mistake marks this out more emphatically and is unlikely to be something a British English speaking candidate would do, where a drop in pitch is more likely for an apology. This different prosodic system and the lack of repetitive wordings, as in the example above, perhaps make the structure harder to follow. Where there is stress, it serves to draw attention to the candidate's error, amplifying any mistakes. Again, the same strategies are used, but show differences when looked at contextually that can account for how candidates sound different (and see 4.5.3 above).

4.6.3.3   Reference and Cohesion

A crucial means by which we understand talk is to have a shared understanding of the reference words we use in English, such as 'this', 'that', 'here', 'there'. We can use these to refer back on ourselves and avoid repeating the same words over and over, but they can become problematic if their point of reference is too vague. The UKG in the haemochromatosis case tends to refer back to only one item at a time, as Example 4-24 demonstrates. Things can quickly become confusing if referring back to multiple items. Some confusion over cohesion and reference structures are compounded by the nature of this being an observed interaction, rather than one in which all are directly involved. In the '*JK*' case which is also a genetics case (Appendix Example C-34), the candidate begins drawing a diagram of X and Y chromosomes at line 194, '`if i show you here it's x`'. However, because of the 3-part interaction structure, it is difficult to tell where '`here`' is and she was criticised for this confusing sequence in examiner feedback. Moments of interactional disfluency have impact on these important explanation phases.

> **Example 4-24 – Referring to talk**
> ```
> 264 CAN: if we talk a bit about
> 265 CAN: you know haemoˈchromatosis⌉ that y-
> 266 ACT:                          ⌊mmⁿ⌋
> 267 CAN: it is (.)
> 268 CAN: what we call a um i mean it's a genetic condition
> 269 ACT: mmm
> 270 CAN: um which means it is inherited
> 271 ACT: right
> 272 CAN: t- um
> 273 PPP: (0.7)
> 274 CAN: now haemochromatosis
> 275 PPP: (0.7)
> 276 CAN: if i talk about what it does it
> 277 CAN: as you're right you're completely right it is too much
> 278                          blood
> 279 ACT: ok
> ```

4.6.3.4   Metaphor

Metaphor can be a useful conceptual tool for conveying difficult ideas and to make equivalences between things. This may also be done with other imagery, such as similes ('it's like a bag'). We use metaphor all the time in making sense of things which are abstract or difficult to grasp (Lakoff 1987: 303). Indeed medicine, which must incorporate many complex scientific concepts, makes use of metaphor all the time (e.g. Kirklin 2001). In this example, the candidate uses the metaphor of the body as a machine (Example 4-25). As with other features that show the relationships between information and ideas, metaphors are also a regular means of alignment, making talk more informal or light hearted as it makes concepts more familiar.

> **Example 4-25 – Machinery Metaphors**
> ```
> 276 *CAN:   um there's er a problem with (.)
> 277 *CAN:   you know something which
> 278 *PPP:   (0.6)
> 279 *CAN:   regulates the transporting
> 280 *PPP:   (0.5)
> 281 *CAN:   of iron
> 282 *ACT:   right≈
> 283 *CAN:   +≈ um and when that's faulty it just builds up in your
> 284 *CAN:   system and it takes a very long time
> ```

4.6.4   Relationships between people

4.6.4.1   Narrative

Sequential ordering is important to most explanations, especially when there is a complex or large amount of information to be linked. Explanations are often given within structures that show resemblances to 'telling stories', or what linguistics would call 'conversational narratives'. This is evident in the way listeners are directed to make sense of sequences of events and their conclusion. The candidate performing the haemochromatosis explanation has already signalled the start of his explanation; '`haemochromatosis,if I talk a bit about what`

it does' (272). He begins (Example 4-26) abstractly, in the present tense, 'it's too much iron in your blood' (273-4). He then switches to a hypothetical mode with, 'over years if you have very high iron levels', now setting up a temporal structure to this description of symptoms, a kind of 'what if' story to account for risk. He then provides a quick list of possible and sequential symptoms with the listing structure 'it can... ', concluding with an emphatic 'eventually' and a final 'can lead to kind of liver failure' (again using 'kind of' as a mitigating marker, see 4.3.2.1). This isn't the full conclusion though, as he produces a kind of 'coda', a conclusion or summing up. Labov and Waletsky's (1967) classic study of spoken narrative addresses these codas and suggest they link the narrative back to the 'time of telling' and encapsulate some kind of point to it all. Here it links the explanation back to the meaning of this condition for the patient – 'which i understand is what your your father suffered from'

Example 4-26 – Narrative structure / sequential ordering to explanations

Codas signal the end of the narrative, and we can see this is what the RP interprets as he begins to talk about what it was like for his father. 'RPL: yes i mean he they kind of went through that ***** with him i mean initially it was you know diabetes (lines 305-8).This is the most he has said for over a minute. Up until now he has been granting the candidate the floor with lots of minimal 'mmm's and 'yeah's. The end of this explanation phase then is clearly marked through this narrative style coda and the candidate and role-player can have a brief discussion together before moving on.

4.6.4.2 Metacommunication – 'Warning – here comes the medicine bit'

We've already talked about metacommunication in Linguistic terms 6 as a means of signalling to a listener how it is you want your talk to be received – it signposts where talk is going but also gives information on attitude and stance to what is being said. So when Jennifer Anniston warns us 'Pay attention, here comes the science' in a shampoo advert, she is 'metacommunicating' not just what's coming next, but also signalling a stance toward the little diagram that pops up as 'science'. It contrives a relationship that Jennifer Anniston and I, the viewer, might have to concentrate extra hard on the tricky, pretend school 'science'.

This strategy doesn't always sound as silly as a shampoo advert. The candidate in this haemochromatosis case (see Example 4-27) does a lot of sense-making metacommunication to warn us about the upcoming 'science bit' during his explanation. He flags up medical terms; `it's what we call...`, a 'we' that means the medical profession, again indicating stance and warning his patient it might be a trickier term. He begins the opening of this explanation by describing what it is they are going to *talk* about; `If we talk a bit about you know haemochromatosis`,`if we talk about what it does`', clearly signalling he will talk about the effects of the illness, distinguishing it from a later explanation on inheritance. This signals the beginning of the explanation phase and the large proportion of the conversational floor he will take. So metacommunicating is a way of signalling what's coming next, but also taking account of the particular listener and how they can relate to the information or stance of the speaker.

Example 4-27 – Extra-metacommunicative work, giving information on how our talk is to be interpreted

```
264 CAN: if we talk a bit about
265 CAN: you know haemo⌈chromatosis⌉ that y-
266 ACT:                ⌊mmm⌋
267 CAN: it is (.)
268 CAN: what we call a um i mean it's a geneti
269 ACT: mmm
270 CAN: um which means it is inherited
271 ACT: right
272 CAN: t- um
273 PPP: (0.7)
274 CAN: now haemochromatosis
275 PPP: (0.7)
276 CAN: if i talk about what it does it
```

```
287 *CAN:   um but over years if you have
288 *CAN:   very high l- iron levels
289 *CAN:   it can cause problems
290 *CAN:   which are the things i was asking you so it can actually cause it
```

However, if the candidate signals this floor-holding explanation phase but doesn't follow this through, it can stand out as incoherent. In 'JL' (see Appendix Example C-32), he appears to indicate he will explain the thyroid tests `should we talk about thyroid then`',

for which the roleplayer explicitly grants him permission, '`yes feel feel free`', but he then continues with further data-gathering. Metacommunication, therefore, as a way of commentating and directing the structure can also make a section of dialogue sound unstructured and 'clunky' if it falls in the incorrect place or is not followed through. Metacommunication runs throughout many of the explanations so can become difficult to analyse as a single category. Successful candidates do the explicit signposting ('In our examination there….') and flag up their definitions ('What we call…'). But most unsuccessful candidates also use these strategies. The difference between failing and passing candidates is in two areas: the signposting that does not fulfil its promise (as in the example above) and the more wide-ranging use of metacommunication to comment on where the speakers are at any moment and any difficulties this poses eg. 'this is rather an unusual request'. There is some tentative evidence that English speakers tend to do more explicit metacommunicating than speakers of other languages. As with the other comments above on different Englishes (here notably English influenced by Indian languages) any speaker's English is dynamic and changing and may display all, some or none of these features.

4.6.4.3 Relating back to and managing patient concerns

Although the explanation phase is quite monologic, candidates use strategies to maintain 'involvement' with the RP. Of course making the information understandable is part of that involvement – through the information giving strategies we have seen above. But there is another involvement strategy many candidates use to maintain a sense of dialogue and sustain alignment within this very monologic phase. They infuse the explanation with the voice of the patient.

The UKG giving the haemochromatosis explanation links this to the prior knowledge and understanding he has already discussed with the patient during data-gathering. So he draws on the RP's own language to give his explanation, aligning to the patient while incorporating the patient's words into his own particular discourse as a joint production (Example 4-28). We saw this linking back to the patient again at the end of the narrative structure with 'which I understand is what your father suffered from'. And he makes a 'metacommunicative' or explicit commentary, pointing back to something that was said earlier in the consultation (Example 4-

Example 4-28 – Linking to patient knowledge

272	*CAN:	if i talk about what it does it
273	*CAN:	as **you're right you're completely right** it is too
274		much iron in your blood
275	*ACT:	ok

Example 4-29 – Linking to earlier talk

287	*CAN:	um but over years if you have
288	*CAN:	very high l- iron levels
289	*CAN:	it can cause problems
290	*CAN:	**which are the things i was asking you**

29). This links the explanation back to the symptom questions asked of the RP, helping make the cognitive links throughout the consultation, as well as justifying why he asked those questions. The structure of definitions, using subordinate clauses, not only gives additional information, they also signal a shift from the medical to the personal – it's what 'we' (the medical profession) call a genetic condition 'which means' (for you the RP) it has some inheritance risk.

In explanations, the candidate has to manage the patient's uncertainty and anxiety. Expressing risk, responding to anxieties, showing one's own uncertainty, being persuasive all require conveying information other than in an assertive mode. This is done generally through modal verbs in English such as 'can', should' 'might' etc. which also act as softeners to show respect and equalise power relations (see 4.3.2.1), and through conditional tenses e.g. 'if you were going to have it, you would have got it by now'. In the haemochromatosis example, the repeated 'can' in the list of symptoms both expresses possibility but is also attuned to the patient's likely anxiety.

4.6.5   Summary – Explanations

- Explanations in the CSA fell into four broad categories (see Figure 11), each requiring a somewhat different focus (e.g. relatively more open–ended options versus the candidate needing to be more persuasive). However, all explanations require extended talk, balancing conveying clear information with creating and maintaining alignment with the RP within an overall narrative structure.

- The detailed analysis of a demanding clinical explanation identified 7 key features of a successful explanation (see 4.6.2 and 4.6.3) and represents a useful template for explanation-giving, adaptable to different circumstances.

- Amongst failing candidates (aside from incorrect clinical information) a range of communicative differences, mostly subtle and difficult to analyse in ongoing talk, can lead to explanations appearing to lack coherence and lead to misalignment.

- These communicative differences have been analysed in relation to the English spoken in the Indian sub-continent where they are used systematically (and is the area where most IMG candidates come from). In sum, they lead to difficulties in tracing the line of argument and processing how phrases and clauses link together.

- All candidates display some of the features of successful explanations but the local lack of coherence and explicit tying of ideas together can undermine these successful strategies.

- People's use of language is not fixed and many candidates from the Indian sub-continent will be influenced by local English and display some or none of the features described here.

Chapter 5.
How examiners assess candidates

5.1   Summary

The data used here are based on examiner feedback of video clips. This is a small data base used to corroborate other findings. While the numbers are not sufficient to be widely generalisable across similar settings, the analysis provided an insight into the processes of examiner assessments and how these related to the sociolinguistic analyses.

Examiners widely acknowledge the challenge of the exam for all candidates and some of its difficulties i.e. its relationship to real world consulting, the demands of simulation, the impact of role-players on the consultations, anxiety and the 'Sherlock Holmes' detective work required of some cases.

- There was widespread awareness of and sympathy for the challenges that IMG candidates face in the exam. There was recognition that examiners judged doctor-patient communication by their own standards (including their own styles of talking and interacting) and some recognition of how diversity in patient populations raises questions about what standards can be used in diverse societies.

- There was considerable agreement in rating candidates' performance in the areas that examiners chose to comment on i.e. largely about manner and affect. These shared interpretations of candidate behaviour is based on inter-group subjectivity, drawing on 'common sense' (see Linguistic terms 9) and professional shared meanings of this relatively homogeneous group. However, there was also variety in examiner assessments. Corroborating other recent research, there were individual differences which are likely to be based on individual style and individual noticing of aspects of behaviour and are part of the inherent subjective processes of evaluating interaction (see also Govearts et al 2007 and Yeates et al 2013).

- Examiner feedback showed that candidate 'manner', both spoken and bodily, and attitude were the performance features most noted. The rare comments on clinical competence are largely a function of the selection of cases ie those with no serious medical mistakes (see Chapter 2). But the examiner focus on manner shows how much of the assessment process depends on *how* candidates interact with RPs across all three domains.

- These comments on manner related, largely, to how candidates sounded i.e. 'formulaic', 'clunky' etc.. Although no comments about candidates' pronunciation indicated that they were considered hard to understand, many of the judgements related to other features of talk such as intonation, pace and rhythm. Quite small behavioural cues in how candidates sounded or their bodily conduct led to general comments on manner and have the potential for larger consequences in terms of assessment. These small differences are the most culturally specific aspects of language (and see Chapter 4).

- These evaluations of how people present themselves are of necessity subjective, reflecting prevailing norms and 'how we would do it' (see Chapter 4 and below). Objective sets of descriptors cannot remove subjectivity (Yeates et al 2013).

- The descriptions of manner and affect used general, informal, evaluative language which provides no specific advice to candidates on how to improve their performance. There is a strong case for introducing an analytic language which has explanatory power for registrars, trainers and examiners.

- Judgements of clinical management were regularly conflated with assessment of manner and affect. Comments on clinical management and interactional management were not often distinguishable.

- Where candidate styles of talk and interaction differ from those of the examiners and role-players (see Chapter 4), such differences have the potential for affecting examiner judgements of the quality and adequacy of candidate performance.

5.2 Introduction

The RCGP has a long tradition of designing a defensible exam in terms of range of cases, criteria etc. and in training examiners (see Appendix A). But precisely how examiners make integrated and global marking judgements about candidates is something of a 'black box' in terms of the particular features of a candidate's performance (Kogan et al 2009). In other words, the surface decisions can be readily seen but the inner workings of how these judgements come to be produced are largely hidden. In attempts to peer inside the 'box,' rather than using questionnaires or research interviews, examiners were asked to react to segments of video recorded cases. These video feedback sessions alerted the researchers to what is noticed but not usually consciously seen or heard in interactions and allowed them to focus on and slow down the interpretive processes so that they are susceptible to analysis. This is precisely the same approach as our linguistic micro-analysis in Chapter 4 revealing those hidden aspects of communication which we perform unconsciously and take for granted in our day-to-day interactions, but which play a centrally important role in how we form our social relationships through talk. Analysing the language of the consultations closely and gaining insights from examiners at this local level, helps us to compare the findings of micro-level performance with the interpretive processes used to make assessments. This was possible because of the very open and co-operative way in which examiners took part in these feedback sessions.

Methods

The data drawn on for this section were: four examiner feedback sessions (the core data for this section). Four feedback sessions were set up lasting between $1 - 2.5$ hours. 20 Short clips were taken from examples across the range both from the 40 case data set and a purposive sample from the 198 video recorded consultations. Examiners were asked to respond to all aspects of the consultation except for clear clinical errors or omissions. They were given the following instructions:

'None of these clips are presented with pass/fail or with a borderline standard setting. Rather than focussing on what grade you would give the candidate, we would instead like you focus on what it is about the candidate's performance that you evaluate as good or poor and what particular aspects of their consultation lead you to this conclusion'.

Since we were only showing clips and not the whole case, examiners inevitably had to focus on moments of interaction, rather than work from an impression of the overall case and whether all the descriptors were fulfilled. This data gathering also meant CSA examiners had an opportunity to be more directly involved with the research. As a research team, these discussions allowed us to clarify more opaque comments about a candidate's communicative performance, such as 'clunky' or 'cold' with greater detail and fine-grained analysis of moments that contribute to these conclusions.

The four sessions were audio recorded and three were transcribed and detailed notes taken from one.

These data were analysed jointly by AS and CR using interpretive thematic analysis. Through joint iterative reviewing of the data, the themes discussed below were identified through a constant comparative analysis (Lingard et al 2008). As well as the analysis given below, these data helped to inform our linguistic analysis and the communicative features of the exam.

In addition, the following ethnographic data also informed the analysis in this chapter: the feedback statements written about candidates who received low marks for a case and/or were categorised as borderline, comments from the KTP advisory group of examiners on low scoring candidates and informal comments from examiners at the exam, conferences, core group meetings and other more informal meetings.

The most common responses to the video recorded cases centred on: (i) evaluative judgements of candidate communication, which included spoken manner, non-verbal communication and interpersonal skills (candidates' affective stance and ways of relating to the actor/patients, including cultural aspects). (ii) evaluative judgements of clinical management (iii) comments on the challenge of the exam. The final part of this chapter comments on the implications of examiners' assessment of candidate performance. It is important to note that the feedback sessions were based on reactions to short segments of video, rather than on formally marking the whole 'case', and that the examiners knew that the researchers running the sessions had social science rather than clinical backgrounds. Both these factors may have influenced what examiners' focussed on in these sessions and help to account for the rare comments on clinical competence since candidates who made clinical errors were screened out. However, examiners were not steered to consider any of the three areas which were most frequently mentioned in their video feedback.

It is not at all surprising that examiners should judge candidates in what are generally acknowledged as common sense ways (see Linguistic terms 9). We evaluate people's talk and body language all the time and comment on how they come across. These taken for granted

evaluations are prompted by widely circulating discourses in the media, the training world etc. These evaluations are also made by patients when they come to see their GP, and part of the function of the interpersonal skills domain in the CSA marking schedule is for the examiner to evaluate the likely effect of the candidate's ability to communicate professionally with a patient (RPs play no part in the marking of the CSA). The difficulty with these judgements being used to assess simulated high stakes licensing exams, is that (i) they can have a disproportionate effect on the assessment of overall clinical performance because of the intense noticing of how candidates present themselves (ii) they are assumed to be universal which raises questions about their standing in an increasingly globalised world and, underlying these first two, (iii) they reflect a set of ideas or schema in a particular historical period, which appearing to provide the answer, and can have great power. For example, Roy Porter discusses the medieval humours as '(a) neat schema with unlimited explanatory scope' (Porter 2003: 47) or Victorian phrenologists used an individual's physiognomy – people's faces or shape of their heads – to read off their internal state and personality. 'Showing empathy' and 'body language' are two of the current schema used to judge the inner person.

Linguistic terms 9 'Common Sense' judgements

'Common sense' is a term we often use without thinking about it. It would seem to mean the ability to understand things in a way that is shared by all people. Of course, shared understanding varies according to different cultural norms, so one community's 'common sense' may be rather different to another's. Much as we addressed with language variation (see Linguistic terms 8), 'common sense' implies a particular norm or 'standard' way of doing things that seems 'unmarked' compared to the non-standard. The obviousness and unconscious process in employing 'common sense' may not be so obvious to those outside the particular community. While we can't escape using our 'common sense' judgements about people on a day to day basis, we can draw attention to this process in better understanding how we make decisions.

While Chapter 4 showed how interaction depends upon moment to moment processing of meaning, this chapter unpacks how this process of meaning is assessed. Interaction is necessarily experienced subjectively, as a set of individual perceptions and interpretations. But it is also 'intersubjective'. This describes the way in which 'common sense' shared meanings are used by people to interpret what is going on. Interaction is always subjective and intersubjective and these terms need to be decoupled from any negative connotation, although they are often used in this way. So 'subjective bias' for example is not a useful term to employ in considering the CSA. What is important is to understand the subjectivity and inter-subjectivity that goes on in consultations and their assessment, then consider what this means for the design and processes of the exam (and see Govaerts et al 2007 and Yeates et al 2013 whose research critiques the notion of objectivity in assessing consultations.)

5.3 Evaluative judgements of communication

Judgements of spoken manner and bodily conduct were the areas most frequently discussed. They are dealt with separately from the section of attitude below because both speech delivery and bodily conduct are omnipresent in interaction and do much more work than only convey attitude and feeling.

5.3.1 Spoken manner

In most cases, there were judgements of candidates' manner and these were the first aspects of the consultation commented on. In all cases where candidates' style of speaking differed from what was called 'the BBC radio 4' style, there were numerous comments on pace, volume, intonation (prosody). There were relatively few comments on vocabulary/jargon. And none of the comments about 'accent' indicated that candidate pronunciation, although perceived as different or foreign, prevented the candidate from being understood.

It was typical for 'non-BBC radio 4' ways of speaking, particularly intonation and pace, to lead to judgements of speaker intention or attitude e.g. 'same level of voice all the way through – no hint of empathy'; 'empathic statements didn't sound empathic' versus 'I like the accent (Scottish) sounds empathic', 'did soften his voice and show empathy', 'come back to see *me* – he took responsibility'. This shows how important prosody is in making evaluations about IPS (see Chapter 4 on formulaic examples and Appendix Example D-1 on a Scottish candidate and Example D-2 on 'cultural bias').

The majority of these comments referred to IMG candidates. Only when there was an examiner in the room who spoke with the same style as the candidate being viewed (a Scottish candidate) was this judgement challenged. Despite these reactions, a few examiners commented that: 'such matters had to be put aside' or 'were not noticed at all'. By contrast, white British candidates' manner was only remarked on if there were regional differences or disfluencies because, as examiners remarked, when candidate and examiner share similar ways of speaking: 'you actually don't hear it', 'you don't particularly notice it'. This suggests strongly that instructing examiners not to notice varieties in speech cannot be readily followed.

The descriptions of what was perceived as poor style used general evaluative vocabulary e.g. muddled, rambling, muttering. 'His English was clumsy in a number of places. He would ramble on in a muffled way... and then end with a slower question in a louder, muttering way, so a mixture of acknowledging the hearing loss but then rambling along in the middle'. It was also commented that 'people' (examiners and trainers) say 'we want them to speak slickly not 'clunkily' by which as analysts we inferred that examiners were talking about how information is delivered as well as candidates' overall style of talking. Given the importance of communication in the CSA, there is a strong case for developing an analytic language for discussing these issues.

Given the analysis in Chapter 4 on the way candidates sound when they are delivering CSA-

type phrases (for all candidates, as discussed above, use very similar phrases, oriented to the requirements of the exam), it seems striking that the prosody of an utterance can make a substantial difference as to whether it is perceived as formulaic and recipe-like, or fluent and sincere. Similarly, a noticed pause, an odd rising tone, the difference between 'how do you feel' and 'how do you feel about that?' or the inclusion of 'actually' were given large consequences i.e. a candidate could be judged as showing sensitivity and sincerity or not. So, repeated small interactional differences could feed into large and significant judgements. This phenomenon was already identified in interactions between RP patients and candidates (above). These small differences affect both the direction of the interaction and how both sides evaluate each other and so also feed into examiner assessments of how the interaction is managed and how relations between the RP and candidate are built (or not).

5.3.2   Bodily conduct

This is dealt with separately from talk only because the feedback comments tended to uphold this separation. Interactional research has shown that meaning is conveyed through the integration of talk and bodily conduct. There was an understandable assumption, given the current discourses about non-verbal communication, that the body 'talks' in an unambiguous and self-evidently significant way, separate from language – summed up in the misleading term 'body language'.

There was a high rate of comments about bodily conduct. It was often the first feature commented on (as part of the early judgements of candidate manner) and these comments seem to privilege bodily conduct over talk in terms of ascribing candidate feelings/ attitudes/ competence /trustworthiness. For example, remarks on candidate behaviour implied that there is a direct equation between bodily conduct and attitude e.g. a direct look = direct talking; open posture = open and non-judgemental attitude and that it is unproblematic to read off from bodily conduct such evaluations as: 'uncomfortable', 'self-effacing', 'defensive', '(un)empathetic', 'distant'. It was also assumed that mirroring the other's bodily conduct was inherently good. The problem with making these evaluations is not that there are no social meanings attached to how people use their bodies but that these cannot be taken out of context from everything else going on and attributed an overall meaning e.g. mirroring may have little or no significance or not be appropriate at all or noticed above all other meanings going on.

The most frequent comments were about: eye-contact and bodily movement, particularly hands, posture, proxemics (how close people are to each other) and fingering the paperwork. Throughout the examiner comments, there was great emphasis on 'eye contact' and maintaining eye contact (even though the function of gaze is culturally variable and the almost universal use of the computer in the consulting room challenges the importance of eye contact).

The attention given to bodily movement and posture may be at least partially accounted for by research which suggests that ambiguity and confusion in interpreting a speaker's message leads to greater attention paid to bodily movement (Kirch M. S. 1979: p 417). Given the findings above that 'empathetic' statements can be perceived as formulaic (and so giving off a mixed

message), there may also be greater scrutiny of failing candidates' bodily conduct.

These examples from two candidates in one examiner feedback session were representative of examiner comments on the video clips:

Comments about Candidate 1: Positive: Head on one side and wide-eyed, looks empathic; hand on heart – heartfelt – interested in her at a personal level; non-verbally showing empathy and concern; still and quiet so facilitating conversation.

Comments about Candidate 2: Negative: a lot of discomfort, fidgeting, twitching e.g. scratching chin, hands held tightly; he was sitting back as if to say 'it's your problem pal'; he was very upright, sitting forward slightly but not empathising at all; judgements coming through in the body language – he was uncomfortable, that came across in the body language; deals with his discomfort by focussing on paper work.

5.3.3 Affect/stance: emotion, connection, rapport, engagement and their opposites

There were more comments on positive and negative affect than any other aspects of the consultation. Most of these are linked to 'manner' (see above) but many were quite general. This suggests that the generic feedback statements on the CSA exam (see appendix xx) mask a lot of the social evaluation work that goes on in examiner assessment since only a minority of these statements are specifically concerned with the emotional stance of the candidate. These frequent comments bring out into the open the extent to which assessments of others' talk/interaction are based on subjective feelings, even though these are not explicitly elicited and documented in the paper work.

5.3.4 Positive affect

Many comments were broad descriptions e.g. 'connecting, open, caring, curious' and were general inferences, for example: 'Something intangible made me think as a patient I could connect with her – the sort of person I would warm to. I actually quite like this doctor and I'm not sure what it is about her'. Candidates were expected to be involved, be 'genuinely interested' and so candidates using expressions that sounded formulaic were easily rated as not genuine.

There were also some specific candidate strategies identified that were perceived as producing positive affect i.e. the use of what we have called 'prefacing' e.g. 'I don't mean to pry'; the use of depersonalisation e.g. 'sometimes after a loss you can drink more' which was rated highly as an indirect way of trying to elicit face-saving roleplayer feelings/reactions; the use of 'empathy' phrases e.g. imagining how the patient might feel e.g. 'I know this could be very frightening for you', or solidarity statements 'this news must have upset you'. These strategies were rated highly and not seen as formulaic since candidates customised these phrases and located them appropriately.

How examiners assess candidates

5.3.5   Negative affect

Negative comments were, interestingly, more frequent than positive given that we showed equal numbers of passing and failing candidates. Again these were closely associated with manner e.g. 'she was so cheerful but that's just the way she spoke – she was a bit hyper'. As with positive affect, there were broad evaluations: 'discomfort', 'lack of empathy', 'couldn't control the inner voice', 'not engaging', 'not listening', 'not interested enough', 'a bit Olympian'. Sometimes inferences from words and bodily conduct were seen as willful e.g. when examiners ventriloquised candidates' thoughts, imagining what the candidate was thinking about the role-playing patient: 'I'm not going to empathise with you'.

Being formulaic was the most commented on aspect (see also Chapter 4) and IMG 'patient-centred' utterances were less tolerated than others e.g. 'mix of perfunctory and formulaic'; 'does not sound conversational and a good consultation should sound like a professional conversation'; 'these stock phrases do stand out and are used a lot by those not doing very well'; 'They sound formulaic to us'; 'they don't have a natural ability to suit phrases to the person in front of them'; 'can't adapt' (and see 5.2.3.1). Another frequent theme was the lack of listening or engagement – 'needs to be more of a conversation'. 'Not listening' or 'not active listening' were used generally and referred to quite a range of features which linked with criticisms of formulaicness. While some of these could stem from the gap between RP patients' cues to the candidate and the latter's inferential processes (an issue for training) other apparent 'not listening' cases may be the result of exam modelling e.g. shoe horning in patient centred questions or performing to the examiner (as shown above, candidates do more talking in the CSA). e.g. '"Check understanding of the patient" – we think it's good. But when learned on courses it can create a hazard. You have to exhibit behaviour – like the driving test. This doesn't mean checking understanding is not happening in high quality consultations, they might not have to say "can I check your understanding"'; (commenting on a candidate's checking of the role-player patient's understanding again): '"I think it came from his heart not the template – he wanted to show the examiner that he'd picked up that it wasn't going right"; "How do you feel" – I don't think he meant how do you feel. A lot of people have been taught this. It's a technique that might be useful but the way we see it being demonstrated it isn't at all. They don't always listen to the role-player's response to the question.'

Overall, IMG candidates in the cases viewed were more likely to be subject to inferences about general communicative and sometimes clinical ineptitude e.g. 'trying too hard', 'doesn't give the impression she knows what to do', 'she talks so fast' and for general comments on sounding formulaic (see Chapter 4)

5.3.6   Examiners' understanding of IMG challenges

Examiners talked eloquently and sensitively of the challenges faced by IMG candidates. Their understanding was based on a mix of cultural and social explanations. For example, the cultural explanations related to over-compliance and a lack of willingness to confront, 'they don't want to lose face by losing their temper'; 'the schizophrenic patient is another example of the

candidate not wanting to confront the patient'. The social explanations related to the experience of migration and marginalisation: 'They are a foreigner in this country, they're feeling isolated and vulnerable, they don't feel they have the strength to deal with it the way we would.'

There was also recognition that it is easier to bluff, cover mistakes etc. if you can draw on a 'BBC radio 4 style' with which examiners will be comfortable e.g. 'the way she used jargon, she got away with it but if it had been an IMG, this choice of words would have sounded really horrible'.

Many examiners agreed that they judged candidates by the standards of how they, the examiners, would relate to patients (and see 5.4.3.2). But the Scottish example (see Appendix Example D-1) raised awareness that the perceived incongruence between IMG candidates and white patients is only one kind of incongruence and that there are other cultural and social incongruences which are not often recognized. In other words, patients in real consultations may have many socially and ethnically different ways from their doctors of communicating and orientating to illness. But these are ironed out in the exam by standardised RP behavior, which, in terms of style, broadly accords with examiners' styles of communicating (ie a local English style). Arguably, the lack of diversity in communicative style in the exam, amplifies the perceived differences between IMG candidates and RP patients.

There was also wide recognition that it was not always easy to disaggregate communicative competence from either consulting skills or the challenges of the case e.g. sexual health case with a young woman: 'I think it's difficult to pick out his problem with communication and his lack of skill. I mean the result is the same but the cause, we can only speculate on it.'

So, overall the examiners were aware that in going for the mainstream consulting style, there was the potential for other styles not to be affirmed and that the mainstream style would not necessarily work for all patients. This was explicitly raised in a discussion with one of the groups who raised the issue of 'cultural bias':

'Where the cultural bias could come in is our perceptions of how they're performing... because in some ways we are culturally biased. I think we demonstrated that with the Scottish girl quite well... so I think coming in with our pre-conceived ideas of what is good style or use of language or words. I think there is potential (for cultural bias) if we go too far down the IPS as being the only thing we are going to assess. If you mix cases and balance the other areas, you can overcome a lot of that. So not 100% sure is the answer.' (see Appendix Example D-2 and Chapter 9 where the notion of 'cultural bias' is discussed).

This self-awareness of the potential for cultural bias is useful to note, as it shows examiners are aware of the dangers of marking from within a narrow cultural range.

And see Berg et al on possible bias from standardised patients.

5.4   Evaluative judgements of clinical management

5.4.1   Conflation of clinical management assessments with the other two domains

Many of the assessments of clinical management contained aspects which related to interpersonal communication and data gathering. Examiners did not always agree in their assessments of candidate consulting behaviour: for example, some saw a candidate as 'being masterful' and others see her as 'jumping about' so, as one examiner says, 'it may just be a matter of style'. Similarly where a candidate is critiqued for not giving active advice (a clinical matter), she was also defended as being non-judgemental (an IPS matter). And the notion of 'active listening' cuts across all domains (see 5.2.3.3).

It was also difficult to distinguish between aspects of clinical management and interactional management e.g. a 'barrage of questions' could mark a candidate down both interpersonally and in terms of the timing and progression of the consultation (both Data Gathering and Clinical Management as well as Interpersonal skills).

5.4.2   Overall organisation

The great majority of reactions to aspects of clinical management related to three areas of the consultation: overall organisation, timing and progression, and tempo. (These assessments were inferred from short clips only, but triggered evaluations of whole consultations). There were two implicit models of overall organisation conveyed, the 'masterful' and the 'flexible' and both implied a criticism of the formulaic. One examiner made the distinction between formulaic and masterful consulting: 'In masterful consulting, (a doctor) produces a hypothesis and tests it and then if it's not that moves on'. The other model was a 'flexible' one in which the encounter was organised around patients' needs and wants (which also implied a critique of formulaic consulting). Differences among examiners suggested there was a fine line between being disorganised and being flexible, particularly in the complex cases.

Some of the highest praise was directed at consultations where the bones of the consultation model did not stick out through the flesh of the consultation: 'What struck me was that she was not using a model or if she was you can't see it. She actually achieved the task in a really sophisticated way'. It is worth noting here that while 'models' were critiqued, successful candidates do a lot of exam modelling (see Chapter 4) and many examiners agreed that there was an implicit model used in the CSA.

5.4.3   Timing, progression and tempo

The majority of the comments related to timing and progression and included the phasing and direction of the consultation. Weaker candidates were seen as moving too fast, 'he just keeps moving' or too slowly, 'progression was wrong', 'he should be way further on' or too indirectly 'he was beating around the bush' and that weaker IMGs were 'anxious about coming to a

decision too early on'. In the DG phase, candidates were commended for being 'slick' in getting through data gathering rapidly and there were comments that weaker candidates were too slow in this phase (although see Chapter 3 where the quantitative analysis shows that this is not generally the case), and that they were asking too many questions in a script like way. But good candidates were also seen as spending a lot of time 'working out what it's all about'. This also suggests that interpersonal evaluations leak into clinical management so that the impression of the right amount of time may be affected by candidates' manner. Again, here, the pressure of the patient-centred model (eliciting ideas, concerns and expectations – see sections above) seems to bear down more heavily on IMGs than others; to re-work the 'fish out of water' metaphor, it could be said that weak candidates are perceived as doing too much trawling and not enough line fishing (see 5.4 below).

The tempo of the interactions was often the focus of discussion and this related to turn taking e.g. candidates cut off patients because of pressure of time, and to speed of candidate utterances. These were sometimes conflated to give a judgement of overall tempo e.g. 'not much space in the consultation'. However the relationship between overall timing/progression and tempo was not straightforward. Weaker candidates were seen as progressing slowly (see above) but asking too many questions quickly 'they are very conscious of the 10 minutes'. Pausing was generally valued as giving more space but some candidates' pauses were interpreted as deliberate slowness to hide a lack of clinical knowledge. So, it has to be the right kind of pausing – the degree to which a candidate's momentary silence is a good 'space' or a marker of ignorance may depend upon how, cumulatively, the case is seen to be going.

5.5　The challenge of the exam

Together with the comments on manner and affect discussed above, the exam was discussed more than any other subject; in particular the exam as a simulation, exam-modelled behaviour, examiner experience of the exam, the design of cases and preparation and training. These added up to a clear appreciation of the weight of the exam.

5.5.1　The exam as a simulation

5.5.1.1　Differences from the real world

There were many explicit comments on how different the CSA was from the real world. There were queries about whether some candidates would act like that in the real world and whether it is possible to give an honest answer in the exam. Also the real world was seen as more forgiving: 'I find that in real consultations, international patients are more forgiving of me and IMGs find that if they really care, then patients are very forgiving of them.' The suggestion being that 'caring' can be conveyed in ways other than textbook formulations of how to convey empathy. So the exam is a context where professional 'performed' phrases come through more often, and the honest responses are minimised. It is an activity where the professional discourses are institutionalised, and become markers of performance. It is likely that this happens in all exams using an OSCE style design (Wass et al 2002).

5.5.1.2   Acting skills

Examiners articulated the acting element of the exam for the candidate (as well of course as for the RP): e.g. whether the candidate was a good role-player or exaggerating because the patient was hard of hearing; e.g. the candidate had 'acting problems because the case was remote from him'. e.g. 'It's amazing how both of them can demonstrate skills like this in a simulated setting. It says a great deal for the candidate'.

5.5.1.3   Identification of 'saving' and 'sinking' moments

Examiners commented on how RPs actively helped candidates (could momentarily 'save' them) and also at times were unhelpful (could momentarily 'sink' them) e.g. 'the way the RP tells the candidate that he has given her conflicting information is 'quite aggressive'. In line with our findings in Chapter 3 (and Pauline Foreman's 2013 conclusions), examiners did not comment on any correlation between RPs 'sinking' and failing candidates. However, examiners noted aspects of RPs' institutionalised behaviour, that they behaved at times more like RPs than patients: (i) putting demands on candidates that patients usually would not e.g. 'because she's a RP she's getting him to explain things' or highlighting mistakes and so putting candidates on the back foot. But in doing so, role-players were pushing candidates to behave in ways required by the exam and so this could also be seen as helping candidates (ii) behaving in an actorly way e.g. 'How much did the RP lay it on thickly and I wondered at what point she thought the confusion had been resolved' (iii) RP familiarity with the demands of the case e.g. 'the role-player uses the word 'option' which normally we would expect the candidate to use. Has she heard it on previous occasions from candidates?' All these imply RP power. Although this power reversal was not explicitly mentioned by examiners, their remarks suggest that RP power necessarily changes the type of interaction that occurs. From the examiner comments it is not possible to unpick how far RPs react to candidate self-presentation generally, to their specific behaviour at any point, are following instructions from written texts and from calibration or are affected by the habitual demands of repeat playings.

5.5.2   Exam modelled behaviour

Examiners identified several features of the exam which put pressure on all candidates: the interactional design of cases and the resulting impact on time management and the constraints and compulsions of the exam.

5.5.2.1   Design of cases 'what's on the tin?'

While RPs and examiners 'know what is on the tin', there was wide recognition that candidates knew there was likely to be something more than what was presented at the beginning of the case, that they had to detective work from subtle cues – what some examiners called the 'Sherlock Homes' approach: 'things that will jump out and bite you at 9:30 mins. A bit of you will be thinking oh have I got this right, rather than thinking this is a real patient and I'm a real doctor in a real surgery'; 'They're constantly aware that something is going to come up and bite them… Unless they get the hidden agenda, they are going to miss the point – so keep asking

what else are you worried about because it couldn't just be a worried woman with a breast lump. Not just CSA training but the whole medical training'; 'A hoovering system – trying to hoover up everything'; 'It's a scatter gun approach'; to cover a broad canvas 'peppering the consultation with every single systems question' (and see 5.4.3). While examiners acknowledged that 'patients come with several concerns and you need to teach candidates to ask that', our data showed they were aware that the effect of the case design was to promote behaviour which could affect candidates' performance adversely.

This was illustrated in the exam case of an elderly man whose wife has died and who is not eating in a healthy way. This is a good example of where clinical management, IPS and the pressure of the exam cluster together to produce candidate behaviour which is readily critiqued. 'In the real world', there would not be this clustering i.e. the Dr could spend more time on how the patient is coping with bereavement and then go on to diet. But the pressure is on, so the candidate is seen as less sensitive (IPS), jumping about (CM) in order to fit into the 10 minutes because of anxiety about hidden agendas which affect how the DG is done i.e. is this about coping, about diet, is it physical or depression?

There was wide recognition that there are cases 'without a correct answer' and that these cases are often linked to difficulty and, implicitly, to hidden agendas – the secret vasectomy case was the focus of the discussion. Many noted that the case was 'unusual': 'The case is about the journey not the conclusion'; 'Vasectomy case is really about the relationship with the partner'. 'So many others in the room: wife, examiner and then GMC and others watching. Like the old oral exam. Not so much the right answer but how to explain their decision. She is looking for the right answer'. 'I think the case is too difficult. If we can't decide, how's the poor candidate going to do it?' 'It may or may not be too difficult, if we are examining negotiating and IPS then she does quite well. She's floundering because she doesn't know what the right answer is'. It must be noted that these comments referred to complex cases only and that relatively more complex cases were used in examiner feedback than occur overall in the 13 CSA cases selected in any one diet.

Arguably, the design of cases with hidden agendas, which all candidates face, can be linked to comments about weak IMG candidates' and their cultural differences. A generalised anxiety among candidates about such cases can lead to weaker candidates being perceived as doing too much trawling, not being sufficiently focussed. This weakness was given a cultural explanation ie that some candidates came from cultures which were conflict and decision avoiding. An alternative explanation is that in exam preparation candidates are taught to put great effort into being patient-centred and identifying all possible unvoiced agendas. In this case the explanation is less about cultural differences and more about the social imperative to play the exam game and pursue all possible hidden agendas. As a result, they may sound formulaic and 'beating around the bush' which feeds into negative evaluations about not progressing the case efficiently. Examiner comments on case design need to be taken up in the e-learning materials, so that less experienced candidates can practise identifying what is on the case 'tin'.

5.5.2.2   Time management

Time management (and see 5.3.3) and the effects of the design of the exam on candidates' behaviour was frequently commented on: 'well in the real world the bell doesn't matter'; 'candidates often delay more before making a decision or a diagnosis – spend a long time on data gathering'. '(He) doesn't go with anaemia explanation because it may be too early for the exam setting. So goes back to data gathering. So slow progress'; 'because it's the CSA going round and round until the 10 minutes is up'. So the OSCE style design of the case produces conditions that in turn produce behaviour that is negatively evaluated. Good or poor exam technique may mask the quality of candidates' performance in real consultations. These and earlier comments on the organisation of the consultation suggest that both candidates and examiners orientate to the CSA as an assessment of whole cases rather than to the specific skills of the three domains.

5.5.2.3   Constraints/ compulsions

Examiners felt that there were both constraints and compulsions related to what to say (see manner and affect sections above): 'Some statements or questions are much more loaded because they are in an exam and so harder to manage' e.g. 'if the patient asks "is it serious?" in an exam this has much more weight for the candidate. She has to pin her colours to the mast, and she may be worried that either response could be considered the wrong one by the examiner'. Although a few candidates are willing to express ignorance of a condition or its management, examiners recognised the potential dangers of this e.g. 'Being happy to say "I don't know" in a real exam isn't so easy'. Similarly, examiners remarked that candidates feel compelled to say certain things because they are in the exam: 'saying "how do you feel" because they want the examiner to hear it. I don't think that's what the average GP would say – instead "does that make sense", "is that reasonable", "How do you…"'. In other words customising the patient-centred model (as successful candidates do, Chapter 4).

5.5.3   Examiner experience

In relating to their own understanding of the exam and being an examiner, many drew on their own experiences.

5.5.3.1   Fear and anxiety

They related the terror of the exam to their own experience in the calibration exercises at the start of the examining day, when they role played the case in front of their fellow examiners on that case.

'The exam is a terrifying experience so you fall back on the lowest common denominator – which is the model – so you revert to the language of the model'; 'here they've got to get through it – the nerves of the process'. In calibration, examiners said they were aware of being watched and yet they were in the advantageous position of knowing the case whereas candidates: 'always say "did I miss something?" That's what causes them the most concern'; 'They're constantly searching almost falsely'. 'I couldn't remember if I had done/said something', 'I was

very conscious of 6 sets of eyes'. While some examiners commented that they were used to being watched, others said that there were much higher stakes and so a higher level of performance anxiety.

5.5.3.2 'One of us' phenomenon

As mentioned above, examiners acknowledged they would judge the candidate's performance against their own: 'you do kind of compare it with what you would (do).' 'A lot of the time I am comparing them to me and what I'm used to'. There were also several occasions when specific candidate behaviour was responded to with 'I would do that'. But there was also an awareness that the standard could be too high if they were comparing candidates with themselves: 'This is a registrar – what standard are we pitching this at? Our own standard, 25 years later?' (and see 6.2.4 on subjectivity). This comparative approach contributes to the variability in assessor rating (Yeates et al 2013).

5.5.4   Exam preparation/training

There were criticisms of preparation and training for the CSA. Training was seen as too rigid: 'We're seeing the effect of structured and inflexible teaching. This may be making them worse consultants – maybe they start off better.' 'On the video on the shared management plan, there is something on giving options so they all think they have to talk about options but shared management plans are not necessarily about this'. 'Trainers as well as courses may also be a problem because they train for COTS and expect to hear 3 options'; 'It depends on how you teach the model – you need to work out what your bio-medical understanding of the patient is – not a model where you say you've got to ask these questions. You can do it more conversationally'. However, aspects of the preparation for the exam were deemed useful. While some examiners critiqued the preparation: e.g. 'trainers plan for every minute of the case', others countered that 'time management is necessary in a busy surgery i.e. "I'll do the medicine bit and rule out all the causes of blood loss."' (see also criticisms of preparation courses from Kanchandani (2011) and Appendix Example D-3).

5.6   Making inferences and judgements

Previous sections have analysed the content of examiner feedback. This section comments on the implications of these processes. There was very little difference between how these judgements were made. The great majority were, understandably, made rapidly, correlating specific and often small performance features with broad judgements of competence. Most people take for granted the inferential processes which allow them to make these judgements and while there is some recognition that such judgements can only be speculative, most evaluation is based on the myriad cues in talk which are not definitive but suggestive. They are a nudge to the inferential process rather than a transparent window into inner thoughts, feelings, knowledge and intentions (Roberts 2000:115). This helps to account for the individual differences in examiner reactions to the video clips (see below and Govaerts et al 2007 whose research shows that assessors are more sensitive to context cues and make more inferences).

How examiners assess candidates

The inferences we make in talk are largely the result of where we grew up and were socialised into particular ways of talking and interacting, and how subsequently we adapted these ways as we communicated increasingly in the institutionalised networks of our profession. So these taken for granted processes account for the high degree of consistency overall in the judgements of candidates, since the great majority of examiners share ways of talking and interacting.

It is no surprise that there can be both high levels of consistency within a group and yet some individual variation since the cues that trigger the inferential process are multi-channelled and complex. At any moment, words, intonation, posture, voice quality, gesture, speech tempo and other cues will be co-ordinated together to convey meaning and be interpreted by the listener. The examiner, as the overhearer, may notice some of these aspects, others will be processed subliminally, and others not processed at all. Much of this is not susceptible to explicit teaching and is the result of socialisation. Both group and individual evaluation of other's talk and interaction are, therefore, subjective. The rationale for 13 different examiner assessments in the CSA is to address individual subjectivity and bring it into line with a group norm. The data from the feedback sessions show both the high level of agreement in the assessment of candidates in the particular areas examiners chose to focus on but also that the same behaviour can be interpreted quite differently.

5.6.1   Shared inferences and agreement

The analysis in 5.3, 5.4 and 5.5 shows high levels of agreement within each group of examiners in rating candidates' performance. Much of the same language was used in doing this e.g. 'connecting', 'formulaic', '(lack of) empathy'. Most of this language described broad categories of emotion or competence but there were also particular features of talk and bodily conduct which were noticed, e.g. 'low tone', mirroring' and aspects of posture. This suggests that where categorisation is available, then features are more likely to be noticed. But also that these categories are developed around features that are more noticeable, in a chicken and egg way. What is noticed is often what examiners bring to the exam from lay knowledge about interaction or generalisations from psychological experiments (usually of informal interaction) which are circulated in textbooks and a range of frameworks of meaning based on social and cognitive factors (Yeates et al 2013).

This shared language also expresses shared assumptions about what talk and interaction convey, e.g. that mirroring your speaking partner necessarily implies 'rapport'. These shared interpretations indicate the intersubjectivity within the group – that is the 'commonsense' and professional shared meanings used to agree on what is going in the consultation. They share a definition of the situation (Seale 2004).

Virtually all the language used in the feedback was general, descriptive and what has been called narrative language eg 'uncomfortable', 'connecting', rather than analytic. While such language expresses a professional's feel for an observed consultation, in terms of feedback to failed candidates and training/preparation for the CSA, such narrative language offers little specific advice on which candidates can act.

These global impressions of candidates also appear to corroborate Yeates et al's research (2013) which suggests that assessors make a general global impression and then have to fit this impression into the boxes on the form. In our own research, only segments were shown and reactions rather than formal assessments were elicited. However, it raises questions about the use and reliance of lists of criteria as indicators of reliability and objectivity, if examiners in all such OSCE assessments tend to judge globally first and then shoe-horn this overall judgement into fixed criteria.

5.6.2   Individual differences and variation

Within this broadly shared way of interpreting talk and interaction, there are individual differences and such variety in judgements has been widely reported (Boulet et al 2002, Alves de Lima 2011, Yeates et al 2013). There were, unremarkably, frequent individual differences in how to read certain linguistic and bodily features and the candidates' stance: e.g. the same candidate was commended for mirroring by some and criticised for being very uncomfortable by others; e.g. being 'still and quiet' was read as facilitating conversation by some and not being responsive enough by others; e.g. there were differences over what constitutes 'judgemental' – one examiner's approving 'non-judgemental' is another's 'need to wade in and get more involved' and there were disagreements over whether in some cases its right to take a stance which could be considered judgemental. For example, in the haemochromatosis case, the candidate asks what the RP learnt from the information about the condition in the hospital letter. One examiner inferred that this means 'The candidate does not have a clue, so "you tell me and I can feed off that"'; while another thought this was an excellent technique for obscure conditions. There were also some differences between these groups of examiners, the project advisory group's judgements and the anonymous examiners grading the same consultations. (see Appendix Example D-1 on the discussion of a candidate). Such variety stems both from variety in noticing i.e. some aspects of performance are seen as more important than others and variety in interpretation of what is noticed.

The paradox of largely shared norms and conventions but also individual differences between examiners can be explained as follows: (i) Individual noticing: some individuals will pick out a feature (one obvious but not necessarily consequential bit of conduct) and, extrapolating from this, will make a general evaluation. Others will tune into these features as part of an integrated and co-ordinated system of talk and bodily movement and will make an interpretation based on all the other meaning making that is going on. (ii) subtle differences in talk and interaction will be judged on an individual assessor's personal style e.g. what effect a pause may have. These differences are similar to those categorized by Yeates et al (2013) (iii) However, very obvious differences from conventionalised styles of speaking, or the cumulative effect of many small but different features will stand out to a group, who largely share the same communicative style, and these differences are likely to be judged negatively (e.g. as formulaic, as unclear etc.) although one or two features of talk may still be assessed positively.

Chapter 6. The Inter-personal skills (IPS) domain

6.1   Summary

- This research raises questions about (i) how far IPS can be effectively and fairly assessed in standardised OSCE exams and (ii) the unwittingly intense focus on IPS in the CSA. This second question is even more pressing in the light of the first question.

- Assessing interpersonal effectiveness needs to be reviewed in terms of the current models and assumptions about patient-centred styles of consulting, the relatively homogeneous notion of what counts as good IPS (see Chapter 5), the CSA 'linguistic fingerprint' (see Chapter 3) and the cumulative effect of small differences in self-presentation (see Chapter 4). Some current literature and our analysis suggests that these assumptions may not be appropriate for an increasingly diverse patient population where different styles of consulting may be more acceptable (see Chapter 7).

- Part of such a review should raise questions about what is taken for granted as explicitly assessable, on the one hand eg 'empathy', and, on the other hand, what is being implicitly assessed as IPS but not recorded as such.

- In particular the taken for granted assumption that states such as 'rapport' and 'empathy' can be judged by observation and explicitly and fairly marked in the exam (see Chapter 4) needs to be questioned. While communication is a vital component of GP consulting, it is not possible to objectively judge some aspects of IPS i.e. empathy, rapport or sincerity from the outside. Empathy is an inner state, experienced (or not) only by someone to whom it is directed.

- And, by contrast, the importance of many small but cumulatively significant differences in how meanings are conveyed needs to be recognised. These are the most culturally specific aspects of language and attempts to repair interactional problems may exacerbate the difficulties. These feed into general assessments of interpersonal effectiveness (as Chapter 5 has illustrated). The inevitably subjective and culturally specific nature of IPS evaluations raises questions about the intense focus on IPS and the standards by which this domain is judged in any OSCE type exam.

- These many small differences in manner and affect occur throughout each role-played case, so that judgements of interpersonal effectiveness are likely to be made across all three domains of the CSA marking schedule (see the micro-analysis in Chapter 4).

- As well as this micro-analysis, an analysis of marking statements shows that the IPS marking statements for the cases leak into both DG and CM domains. So IPS is frequently marked more than once in a case.

- So Interpersonal Skills are both explicitly and implicitly assessed (see Chapter 5) throughout the consultation, giving an intense focus on this linguistically and culturally demanding aspect of the exam.

- A hypothetical test was run on the effects of removing the IPS domain from the assessment design. This test showed that its removal would affect the marks of both IMG and British candidates but only marginally i.e. 4% more IMG candidates would pass and 2% more of British candidates would fail.(However this does not account for the implicit assessment of IPS throughout all cases.)

6.2   Introduction

The analysis of the 40 cases and of the examiner feedback made the IPS domain a particular focus of the research. Managing the patient-centred model depends crucially on how manner and attitude are judged. The concentration on interpersonal communication reflects the major discursive shift in consultation skills in the UK over the last 40 years, as promoted by successive governments, the Department of Health and the General Medical Council. Patient centredness is one of the principles of assessment in the CSA. The data analysis in Chapters 3, 4 and 5 identifies many of the different elements of patient-centredness and IPS and how they are assessed. These are drawn on in this section and are supplemented by an analysis of the CSA paper work on the three domains of the exam. The IPS domain includes communication skills, as the examiner feedback comments and CSA paperwork demonstrate, and how those skills are used to deliver clinical messages and work with the patient to both understand the patient and engage the patient in a shared understanding and purpose.

6.2.1   Discursive shift to patient-centredness

The major discursive shift from doctor-centred to patient-centred consulting over the last 40 years (Balint 1970, RCGP 1972, Stewart 2001) has brought with it a shift in how to relate and communicate with patients. It is argued that interpersonal skills should be rated alongside and as an equal partner with the bio-psychosocial approach (Howie et al 2004). Interpersonal skills and communication skills generally, and 'clinical empathy' (Halpern 2003) in particular, appear in virtually all models of consultation and are taught as a matter of course (Bonvicini et al 2009, Kurtz and Silverman 1996, Hojat et al 2002, Neighbour 1987).

This paradigm change to 'patient-centredness' – a movement from 'detached concern' (Fox and Lief 1963) to empathy and rapport – rarely acknowledges that it is a product of its time rather than any absolute set of permanent standards. For example, the standards of, say, the 50s and what constituted a 'bedside manner' are, in many ways, as culturally different as the taken for granted 'cultural differences' perceived in some overseas trained doctors of today. Stepping outside the current discourses, raises questions about how patient-centredness is best assessed in contemporary society.

Relating to patients in a patient-centred way has been construed as a shift from cool to warm in 'the larger discourse of medicine as it abandons the objectifying and detached clinical gaze for something warmer – a new professional intimacy' (Bleakley 2003: 187). This shift is

supported by the wider discursive shift to a more feminised discourse of caring and co-operation (Cameron 2000) and to, for example, more conversational styles (Fairclough 1992, Carter 2004). The power of these discourses is apparent in all assessments of consulting skills where women regularly outperform men and of course the CSA is no exception. The upshot from these assessments is that women are judged as better relaters and communicators than men and that this aspect of consulting dominates much of the assessment process. It could be said that a 'feminised' consulting style has been in vogue over the last few decades. However, other demographic changes such as the increased diversity in the patient population and among health professionals suggests that the conditions under which this feminised style developed may be changing and that standardised communication and interpersonal skills are in need of review.

As a high proportion of all male and female candidates pass the exam, the differential between male and female success is not particularly remarked on. But when one social group has a high failure rate, as with the IMG candidates, who might be assumed to find the IPS domain more challenging due to language and cultural differences, then questions need to be asked about the focus on IPS. This is not to argue against the value of clinicians ensuring they understand their patients' viewpoint, or making sure they explain medical issues and engage patients in the management of their medical problems, but to ensure that for the purposes of the CSA examination that the IPS domain is not being given additional weighting due to its presence as a named domain in the marking schedule and the implicit evaluation of interpersonal effectiveness throughout the consultation.

6.3   Some critiques of the IPS domain

6.3.1   Assessing IPS

There is a continual quest for IPS definitions, categorisation of skills, instruments for training and assessing, where particular aspects of affective behaviour can be correlated to checklists of skills or to patient satisfaction audits. This suggests it remains an untamed horse. The overarching problem is that in the research and assessment tools IPS/communication skills are conceived as universal and not relative to particular social groups. So, one particular way of showing interpersonal effectiveness and patient-centredness comes to be the norm, as is evidenced in the similarity of textbooks in this area. This is an issue for all settings where there are OSCE type examinations and can also affect work based assessment in any medical training programme.

Howie at al 2004, in a wide ranging discussion of the problem of measuring and assessing interpersonal effectiveness, confirm the difficulty of 'trying to operationalise patient-centredness through analysis of video-taped consultations', and argue for more emphasis to be put on patient satisfaction (and see Campion et al 2002 and McKinstry 2000 on the difficulty of assessing patient-centredness). By contrast, others are critical of relying on patient satisfaction questionnaires for assessing interpersonal effectiveness.

The literature raises substantial questions about whether interpersonal and empathetic communication can be reliably identified, taught and assessed in OSCE-style examinations (see Appendix E-1 Assessing Empathy). Here, the IPS 'horse' is so hobbled and constrained that it cannot be itself. For example, the very fact of training in empathetic expressions can produce perceived formulaic performances which are seen as lacking in empathy (see Chapter 4). Indeed 'trained empathy' is something of an oxymoron. Bleakely (2003) argues that medical education needs to address the deeper level where doctors reflect on inner and unacknowledged drives. However, this does not address the problem of how IPS should be assessed in a clinical examination.

More profoundly, is it possible to assess empathy at all? (Hojat et al 2004 and see Appendix E-1). Only those who are on the receiving end of intended empathy can say how it felt (Seale et al 2007). A smiling young woman might seem empathetic to an observer but phoney or over familiar to a patient. Empathy is what the patient/Dr respond to in each other and cannot be readily judged from outside (and of course in the CSA it is simulated empathy that is judged by a third party). Again, inner states are read off from outward (simulated) behaviour (as in Chapters 4 and 5).

6.3.2   IPS and the CSA

In producing indicators and criteria for marking, there is the additional problem of how to distinguish patient-centredness in terms of appropriate data gathering and management, eg exploring the impact of an illness on a patients' life, and the overall manner in which the consultation is conducted ie the relationship building, 'rapport' and 'empathy'. It is clear from examiner feedback that overall manner is what is noticed and assessed above other (non-biomedical) considerations in reaction to short clips. And as well 'empathy' and 'rapport' occurring quite frequently in the specific case marking criteria, this feedback suggests that candidates' self-presentation and how they manage the social relations of the consultation is crucial to their relative success. Even where such terms such as empathy and rapport do not explicitly appear, as is the case in the CSA generic indicators, there are significant implicit criteria throughout the consultation eg the indicator 'responds with interest to needs and concerns' often leads to the expression 'I understand how you feel' which, if spoken with particular intonation (see Chapter 4) was judged as 'unempathic' by examiners. The negative generic indicators for the IPS domain also touch on such matters eg 'lacks warmth', 'appears patronising'. Together with the examiner comments in Chapter 5, this suggests that it is when things go wrong that these more general IPS assessments of manner and affect come into prominence. So, checklists on interpersonal criteria only partially indicate how IPS is actually assessed and this adds to the more wide-ranging discussion above on the difficulties of conceptualising and assessing IPS.

6.3.3   Empathy for all?

Chapters 3, 4 and 5 have shown that the CSA is performed within a patient-centred/shared decision making model. This is reflected within the relative weighting given to IPS. However

The Inter-personal skills (IPS) domain

some research suggests that not all patients want this model (Howie et al 2004, Ruusvouri 2007). It is also argued that focusing on empathy can undermine what patients may prefer which is autonomy, dignity and expertise (Bouma 2008 and see appendix E-1 on empathy). Some research has also shown that some older patients did not want shared decision making (Howie et al 2004). While some cases in the CSA are designed to acknowledge a more doctor-centred approach eg a case of an acutely ill patient, the generic indicators for all cases in the three domains, assume a patient-centred model.

While particular groups, such as the elderly, may not align to all aspects of the patient-centred model, there is a larger question of ethnic/cultural/language differences. Most research assumes that interpersonal communication skills are universal, and consultation skills research usually ignores sites where health professionals and/or patients are from different cultural/linguistic backgrounds. Where they are included, findings from these sites tend to be anomalous and researchers acknowledge that 'much still has to be done to make performance measurement into a culture–sensitive and equitable science' (Howie et al 2004 and see Frankel 2009). Our own research with linguistically diverse patients strongly suggests that patients orientate less to the widely accepted patient-centred model of sharing decisions and social relationships and more to the clinical aspects of the case (Moss and Roberts 2005) and that what most linguistic minority patients respond to well is a 'looking glass' take on communication text books. For many minority patients, as for Alice in Lewis Carol's classic, 'things go the other way' (Carol 1871) and these patients want the reverse of these text book models (and see also Schouten et al 2006, Blackhall et al 1995)

In complex, morally questionable cases, the goal of a 'warmer professional intimacy' (Bleakely 2003) may not be suitable for two reasons: firstly, warmth, intimacy and empathy are not appropriate when patients and doctors hold very divergent views e.g. a patient who wants a secret vasectomy (*'TJ'* case), or a young teenager who is sleeping around (*'KM 1'* and *'KM 2'* cases). While these cases are in the minority, there is a balancing act between being non-judgemental and engaging more deeply and between withholding affiliation and appearing overly cool or disengaged.

Secondly, the assessment of warmth and empathy in such agonistic cases (see Linguistic terms 2) is subtle and complex and depends crucially on judgements of manner (see Chapter 5 and 6.3.4). Much of this balancing act is done through shared communicative style. In other words, 'pitching' at the right level of formality, shared rhythms and intonation, the routine metaphors we live by etc. are all involvement strategies that are suggestive, in subtle ways, of 'entering the patient's world' which work well when doctors and patients broadly share a communicative style. However, when there is not this sharedness, empathy statements can sound patronising, presumptuous or out of place to patients whose world the doctor is far removed from.

6.3.4　Judgements of manner

Section 6.4 below discusses how IPS cuts across all three domains in the written descriptors, feedback and marking statements. These written criteria are then translated into ways of noticing *how* a candidate performs. These judgements of manner permeate all three domains. For example, the following are specific marking statements from the 40 focal cases and come from all three domains:

1.　Clarification of patient's narrative, ideas/ thoughts, and requests

2.　Recognising implications of a certain position or activity

3.　Making assumptions about patients

4.　Explaining aspects patients need to know about their problem

5.　Acknowledging the patient's point of view, preferences and likely behaviour

6.　Taking into account patient, or other's concerns

7.　Exploring implications of choices

8.　Explaining treatment choices in a way the patient can understand

9.　Explaining the confidential nature of the consultation, and keeping the patient 'safe' e.g. by offering a chaperone for intimate examinations

10.　Avoiding damaging relationships the patient might have with others

11.　Overall sensitivity in the conduct of the consultation, avoiding lecturing or haranguing the patient, and able to deal with most issues without embarrassment or sounding judgemental.

12.　Eliciting social concerns/embarrassment.

These statements require examiners to assess *how* the clinician takes a history, responds to the data gathered, makes a diagnosis, manages the case clinically and explains this to the patient and therefore the overlap with the data gathering and clinical management domain is inevitable. For example, in a genetics case, the ability to 'explain risks to the patient' is in the clinical management domain and 'communicates risk' is in the IPS domain; and the verb 'explore' occurs frequently in both Clinical Management and IPS domains and has undertones of sensitivity and mutuality built into it which depend crucially on manner.

Patient-centredness and interpersonal effectiveness are expected throughout the consultation and our data have shown that, indeed, there are very few cases where candidates do not ask patient-centred questions or acknowledge the RP-patients' contributions (see Chapter 3). The low marks in IPS relate to the manner in which the attempts to be patient-centred are made (see Chapters 4 and 5) and this focus on manner is the object of examiner attention throughout the case.

The focus on manner is particularly hard for those groups where both examiners and RPs sound

'other' to the candidate and the candidate sounds 'other' to the examiner and RP. While the statements mentioned above seem to require that candidates produce appropriate phrases, as Chapter 4 has shown, following these exam requirements may not lead to good marks in IPS. For example, 'take account of patient concerns' may produce explicit questions/statements which sound 'clunky' and in the very act of trying to demonstrate IPS, candidates are perceived to fail to do so. Similarly, many examples from textbooks and examiner comments assume that doctors can display an imagination of the patients' experience in their responses e.g. 'quote imaginatively what the patient might say' (and see O'Grady 2011) or infer how they feel from what they say (see Chapter 4). However, 'recognising the implications' of what the RP-patient has said can sound like intrusive labelling e.g.: 'I know this could be very frightening for you', if the candidate's manner is not viewed as appropriate.

6.3.5   Subjectivity and the amplifier effect

It was recognised by many examiners that they judged candidates on how they would consult themselves (see Chapter 5). As one said, 'How else would you judge empathy?' And these judgements were routinely made on manner, affect and emotion and on quite fine differences e.g. adding in 'actually' or not (see Chapter 5 above). So even small additions or changes to routine phrases can affect assessments.

These judgements were made not so much on the referential meaning of talk i.e. the content of words and grammar of the main message but on those other features which are noticed but not easily recalled (although even in the grammar and choice of vocabulary in talk there is an interpersonal element (see 6.3.6). We have discussed and illustrated several of these noticed but not recalled areas above: prosody – how intonation, rhythm, volume are used; ways of softening and mitigating language (e.g. 'sort of', 'kind of'); modals such as 'might', 'may', 'should'. These more conversational means of communicating help to glue interactions together in a face-saving way eg. 'I think I'd have a word with Mrs X' is a strong suggestion (even an instruction in hierarchical contexts) to the listener to go and talk to Mrs X (and not a statement of intent by the speaker to talk to her; eg. metacommunicating, i.e. commenting on what has been said, may be said, signals as to how it is to be interpreted etc. As Chapters 4 and 5 have shown, when these features are used to promote a conversational encounter, candidates are usually rated highly. So it is not the dictionary meaning of words but how they are used to produce a particular effect and be understood in context; both the 'how' or delivery of talk and the other stuff of talk which helps to minimise uncertainty and to manage the business of social relations. These aspects of talk and interaction are the most culturally different aspect of language and varieties of language and strongly suggest that there is an implicit model of what counts as good interpersonal effectiveness based on educated norms of local English use.

Differences depending on where speakers learnt to speak English not only impact on social evaluations of the speaker, but also the very means to deal with misunderstandings or social discrepancies themselves can amplify the differences. For example, candidates, whether successful or not, had to manage moments of misunderstanding, social awkwardness etc. and

would metacommunicate to get themselves out of trouble. However, attempts at repair work by candidates whose expert English was not learnt in an English speaking country often led to an amplification of the original uncomfortable moment because of how the repair was done (see example 4–12). As discussed in Section 4.4, all candidates have some difficulty in recovering from serious or protracted misunderstandings, but even small misunderstandings (MUs) and misalignments (MAs) between IMGs and RPs were difficult for this group to recover from.

Examiner feedback has also shown that although there is not necessarily consistency in IPS judgements at a particular moment, there is, however, consistency in the way lower IPS scores stack up against certain candidates. This apparent paradox is discussed at the end of Chapter 5. In sum, there are differences in rating IPS/communication skills which are likely to be more disputed because of their subjective nature but these differences are suppressed over the 13 sets of judgements because of the relative homogeneity of examiners' communicative style. Subjectivity is both an individual and a group matter.

Revisiting the analysis of the low IPS rated candidate discussed in 4.3 above, illustrates this point. This candidate passed the case but did not do well in the IPS domain, in the particular case of possible child abuse. And although he passed the CSA, his IPS score was low overall. So, candidates can pass even if they score low marks on IPS but it can pull down overall scores consistently. Analysis of the intonation patterns, particularly the level of pitch and the chunked units, showed that these contributed to him sounding formulaic and the interaction perceived as lacking in rapport. However, when showing small clips like this to examiners, many questioned why this candidate's style was seen as problematic in the first place and the question was then raised that something must have happened elsewhere in the consultation to lead to this evaluation. Although no one single, small moment like this of 'formulaicness' or misalignment would lead to a poorer mark, we would suggest in the case of this candidate there is no obvious stand-out moment where 'rapport' fails. It is the accumulative effect, rather, of the way he delivers involvement strategies like this over the course of the whole case that leads to a general feeling, on the part of the observing examiner, that it doesn't sound quite right. The fact that he does least well in his Interpersonal Skills mark overall would also seem to suggest there is a consistency to the way he stands out as not quite flowing and natural in the IPS domain. This in turn raises several questions (which are addressed in more detail in Chapter 7): (i) Should candidates who sound a little awkward or whose interaction lacks some smoothness attract low IPS marks? (ii) to what extent is the examined patient-centred model forcing some of this lack of flowing, natural talk (with candidates feeling they *have* to insert frequent patient-centred 'empathy notes')? (iii) under different consulting conditions would this candidate's consulting skills be flowing and natural?

The Inter-personal skills (IPS) domain

6.3.6   Empathy in institutional contexts

As with the balancing act between showing understanding but not compliance with patients (see 6.2.2 above), there is a similar balancing act necessary in some cases between patient alignment and (not always compatible) institutional imperatives. e.g. in the '*MS 1*' and '*MS 2*' glucosamine case and in the case '*MC 1*' and '*MC 2*' in which the patient's mother wishes to administer drugs secretly to her schizophrenic son. The notion of 'managing emotion' (Ruusuvuori 2007, Bolton and Boyd 2003) also suggests this tension between patient and institutional/professional stance. It is implied in some of the case marking descriptors but is not always made explicit and may therefore contribute to the use of 'empathy statements' when a more nuanced response is required. If the balancing of alignment and institutional constraints is considered to be an important competence, it needs to be more explicitly addressed in the paperwork and in preparation for the exam.

6.4   The three domains and the CSA paperwork

6.4.1   Analysing the paperwork

The paperwork related to the IPS domain was analysed in relation to the other two domains, Data Gathering and Clinical Management in order to examine the overall structure of the CSA, and the way its content is divided into separate domains for assessment purposes. All 29 separate cases of the chosen 40 case videos (some cases were videos of the same cases as taken by different candidates) were analysed. All IPS statements in the paperwork used to guide examiners in their assessments were reviewed and the extent of overlap between the Data Gathering/ Clinical Management/ Interpersonal Skills domains of cases in the CSA was analysed. The overlap between the three domains has been acknowledged since the CSA was first run in 2007. The psychometric advice on the construction of the three domains acknowledged that they were not 'orthogonal' i.e. measuring separate aspects of the consultation. However, no research had been done to assess whether any one domain was over-assessed. The following paperwork was reviewed:

- Generic marking statements, as divided into the three marking domains of Data Gathering (DG), Clinical Management (CM) and Interpersonal Skills (IPS).

- The actual distribution of case specific marking statements within the case specific marking schedules for the 29 cases. (NB some of these cases appear more than once in the 40 videos used for detailed analysis).

- Background notes on each case

- Supporting information for examiners

6.4.2   Case marking schedules across domains

Each marking statement from each of the marking domains was assessed for its suitability in the domain it had been allocated. The case specific marking domains are each divided into statements that add positively to the grade and statements that would negatively impact on the grade decided for that domain. They are not intended to be summative and examiners are trained to use them as guidance rather than as a checklist. A table was constructed of marking statements that appeared to be allocated into 'another' domain by case writers (see Appendix Table E-2).

This analysis shows that the marking statements most likely to be allocated to another domain group come from the Interpersonal Skills domain. This means that IPS is being considered within the mark given to the DG and CM domains, as well as to the IPS domain. While this happens a little for the other two domains, their contribution is small in comparison with the IPS domains.

- Number of cases where IPS is found in DG = 22

  e.g. 'is alert to sensitivity of the subject, picking up on verbal and non-verbal cues' in case about a sexually transmitted infection.

- Number of cases where IPS is found in CM = 25

  e.g. 'reassures x that y will not be taken into care, based on information provided' in case about child protection

- Number of cases where DG is found in CM = 3

  e.g. 'Explores work-related issues – impact on work, safety at work,? sick note' in case about knee pain

- Number of cases where DG is found in IPS = 9

  e.g. 'Considers effect on driving' in case about Parkinson's Disease

- Number of cases where CM is found in DG = 3

  e.g. 'Recognises implications for future fertility' in case about chlamydia

- Number of cases where CM is found in IPS = 1

  e.g. 'Fails to respond to x's request for help with her weight' in case about overweight teenager

There was no case where there was no reported 'leakage' of marking statements to domains outside their own domain. Cases with only 1 or 2 leaks tended to be the straightforward clinical cases.

The Inter-personal skills (IPS) domain

6.4.3   How far is IPS assessed?

The mapping and review of the paperwork suggests that there is a risk of over-assessing IPS and that this domain is of a different order from the DG and CM domains. The latter two can be seen to be happening in largely distinct phases of the consultation, and certainly the DG phase is fairly easy to recognise and therefore to mark. The marking schedule assumes that these can be distinguished from the IPS domain. However, as Chapters 3, 4 and 5 also show, IPS has a capillary effect on all aspects of the CSA and the domain mapping outlined above shows that IPS, taken across the three domains, is the dominant one. This 'leakage' of IPS across all three domains could lead to the risk of a candidate being implicitly marked twice, or even three times, on Interpersonal Skills within one case, if the other two domains contain a significant IPS content as well. Cases that include more 'emotional' or psycho affective content, such as those requiring negotiating skills or recognition of ethical or moral issues, are more likely to show leakage of IPS into other marking domains.

There are more indicators of IPS than of data gathering and clinical management in the generic indicators: 18 IPS indicators, 10 for CM and 7 for DG. However, there is very little in the background notes and references provided to the examiner for each case on what constitutes IPS. It is the least backed up with expert knowledge, which strongly suggests that it is judged by examiners' subjective assessments of what counts as interpersonally effective. As noted above, there is nothing inherently negative about subjectivity since all interaction depends upon it. However, it is the area of the exam where the examiner assessments are most dependent on people's cultural/linguistic and social experiences as a professional group, with the great majority educated and trained in the UK.

The leakage of IPS into all domains in the paperwork analysis is also well evidenced in the examiner feedback (see Chapter 5) and in the microanalysis in Chapter 4. It is also widely supported by linguistic analysis of the role of the interpersonal in all aspects of interaction. For example, even those aspects of communication which seem the least interpersonal and are about conveying or eliciting information have an interpersonal element (see Explanations in Chapter 4). Recent work in Australia also supports the phenomenon of IPS over-assessment (O'Grady 2011: 320). Interpersonal communication is both explicitly assessed in the IPS domain and implicitly assessed throughout each case.

6.4.4   Effect of removing the IPS domain

As a result of the findings on the case marking schedules, the MRCGP psychometrician, Richard Wakeford, was asked to model the effect of removing the IPS domain from the marking schedule, based on a diet of marks results from January – March 2013 (see Appendix Table E-3). This modelling was done on the assumption that if the IPS domain leaks into other domains, it is being marked elsewhere in the marking schedule and removing it should not affect the validity of the marking schedules.

The overall findings if the IPS domain is removed completely from the calculations:

- The fail rate for IMGs drops by 3% (from 60.9% to 57.7%) on a first attempt.

- This has a greater effect for IMG men than women of about 4% (dropping from 67.3% to 63.3%).

- UK graduates have a slightly higher fail rate of 2% (from 7.6% to 9.7%).

- In total the overall fail rate increases slightly by 0.8% (from 21.4% to 22.2%), reflecting the higher number of UK graduates taking the CSA overall.

- There is overall, no real difference for men.

These figures can be interpreted in several ways and this depends on the extent to which the other evidence from this research is taken into account. While the numbers are statistically significant, they are not remarkable. They suggest that there would be some slight increase in IMG pass rates and the gap between UK graduates pass/fail rate would be marginally decreased. One interpretation is that removing the IPS has only a small effect therefore the IPS domain is not an important factor in the differential success rates. However, this does not take account of the fact that IPS is also being marked in the other two domains of DG and CM, although not classified as such. The effect might have been different if all the IPS descriptors in the other two domains had been taken out of the assessment as well as any judgement of manner made on the consultation.

6.4.5   IPS and standardisation

Reliable assessment implies standardisation. The CSA is standardised in the palette of cases selected and works towards standardisation through the calibration of cases, the sets of descriptors and feedback and marking statements and through examiner and RP training. However, standardisation of interaction is not possible and arguably not helpful outside the scripts given to workers in the service industries (Cook-Gumperz 2001: 123). Once there is any protracted interaction, scripts fall apart and talk that sounds scripted is judged negatively.

IPS is the area where standardisation is most problematic since this is the domain where the interactive and communicative elements of the exam are most carefully scrutinised, where examiner judgements are made based on features of talk processed subliminally and where RP-patients (however carefully trained/calibrated) are not automatons reciting scripts (see weight of exam section below). The goal of reliability within the IPS domain comes to reinforce that there is an implicit 'one best way' of doing it, that there is little room for candidates to do anything except stick to the institutionally idealised mode of patient-centredness and empathy (although a few do not) and some will be better at knowing how to perform that particular style (and actually how not to 'overdo' it) than others (see Chapter 4). The standard questions/statements across the board used by candidates, discussed in Chapters 3 and 4, demonstrate this, although examiners instinctively know that interaction does not work in this standardised way. Some of the complexity can be explained by Brunsson and Johnson's theory (2000) that people operate with dual systems that tend to be decoupled from each other, i.e. that there are formal structures and standards that are not followed in practice. And conversely,

and in addition, in the case of the CSA, that models and standards are more present in the exam than in real life consulting. For example, there were frequent comments in examiner feedback sessions that candidates should perform as they would in day to day consultations whereas this research has shown that there is a widely accepted, if implicit, model of how do to the consultation which defines it in the exam setting (see Chapters 3 and 4).

This chapter raises issues about the current emphasis on IPS and the extent to which aspects of it can be fairly assessed. There is, without question, a level of appropriate behaviour that all GPs would be expected to display, that is largely determined by the GMC as regulator of medical standards and behaviour in 'Good Medical Practice'. To our knowledge the 'Serious concerns' box that would include inappropriate behaviour has never been ticked. On the contrary, in the 40 focal transcribed cases, role-playing patients were treated with respect and the patient-centred stance was widely in evidence. The perceived differences in weaker candidates' performance concerns the manner in which patient-centredness is done and so raises questions about what counts as inter-personal effectiveness, whether it can be standardised and how it is assessed both explicitly and implicitly.

Chapter 7. The weight of the exam and International Medical Graduates

7.1   Summary

- The exam is set within some of the wider themes and assumptions of today's society: a focus on communication and the interpersonal, technologisation and globalisation. While this report suggests that the interpersonal is over-weighted in the CSA, and is difficult to standardise more generally, by contrast the impact of technology and diversity on consulting skills is under-recognised in the CSA.

- The weight of the exam: Many features of the hidden curriculum of the CSA produce its particular weight: (i) The exam requires candidates to manage multiple frames: modelling patient-centredness, reacting in a real and intense social encounter and managing the simulation – both mimicry and a heightened reality (ii) the decontextualised nature of the consultations, in particular the lack of doctor-computer interaction, which contributes to the 'talk heavy' aspect of the CSA (iii) the particular attention given to candidates' 'manner' i.e. how they sound and manage longer stretches of talk.

- While there was no evidence that pronunciation as such was a factor in these ratings (see Chapter 5), the patient-centred modelling requires additional communicative resources to prevent it sounding formulaic (see Chapter 4).

Effect on IMG candidates:

- These features of the 'hidden curriculum' concentrate the gaze on just those aspects of IMG performance which present the greatest challenges. In addition: (i) Simulation presents particular problems to IMG candidates: the differences in communicative style between most IMGs and role-players; the lack of experience among IMGS of role-playing at school or in their undergraduate education (ii) IMGs' regular experience of consulting with BME groups with whom they share a cultural/linguistic background is not assessed in the CSA, although widely reported by these groups (iii) Evidence from other research suggests that IMGs experience anxiety, anger and fear because of the high failure rate of this group which may lead to insecurity about themselves and about their competence in communicating which affects their performance. And research outside the UK suggests that professionals from abroad consider that soft/communication skills are over-valued in their new country whereas their professional expertise is under-valued.

- There was no evidence that RPs consistently contribute to 'sinking/saving' particular groups of candidates. However, they do have power in the exam and very small interactional differences can affect candidate performance.

- Ethnographic information suggests most trainers and IMG registrars considered practical skills developed in ethnically mixed groups was the most effective way to prepare for the CSA.

UK Graduates

- UK graduates possess the linguistic/cultural capital required of the exam. They are also advantaged in that: the diversity cases in the CSA do not assess their cultural/linguistic skills in consultations where they might be fish out of water; they also have considerable experience of simulations and so a feel for the game.

- UK BME candidates do not display the differences in 'manner' of many of the IMG candidates, except for a small number of UK graduates. This latter group may have been schooled abroad and only educated in the UK from undergraduate level (although there are no records to dis/confirm this). This might contribute to an explanation for the failure of a small number of UK BME graduates and needs further enquiry.

7.2   Introduction: the wider context

In the metaphor used to preface this research (see 1.2), some fish feel the weight of the water while others do not. It can be argued that all candidates should feel the weight of the exam, otherwise it is not sufficiently demanding or set at a high enough standard. It is also a truism that individuals will experience this weight to different degrees. The questions addressed in Chapter 7 are: is this weight felt more heavily by one group than another? If so, what are the characteristics of the exam that produce this weight and to what extent are they justified? This chapter discusses the design and conditions of the exam and its implicit model and how these are experienced by all groups of candidates. This discussion is first set, briefly, within a wider frame of the current practices and discourses which are influential in the CSA, and many similar gatekeeping encounters, where face to face interaction is assessed, and which raise questions about standardisation.

Over the last 30 – 50 years there have been five important changes which have affected institutions and professional services and the ways in which medical face to face assessments are made: (i) what is widely termed the 'neo-liberal' focus on the self and soft skills, in which the social and cultural self becomes the centre of attention in assessment, rather than acquired technical expertise (Grugulis and Vincent 2009); (ii) feminisation of institutional talk – aptly summed up as 'emotional housekeeping' (Karpf 2006:283 and see Chapter 6); (iii) the rights and autonomy of individuals (e.g. active citizen, patient power); (iv) technologisation and (v) globalisation. These discourses are so powerful and taken for granted, ingrained from the beginning of UK medical training (see Tomorrow's Doctors 2009) that when put in 'the box' of the exam, it is hard to look outside them.

The first three of these feed directly into the patient-centred model and focus on interpersonal skills discussed in Chapter 6. The latter two also have a profound effect upon interaction in the consultation. The Electronic Patient Record (EPR) affects how doctors and patients relate and interact together (Pierce et al 2008, 2009), with, for example, more gaze on the computer, EPR

tasks being accomplished as the patient talks and the computer's 'voice' taking over at times as the authoritative voice in the room (Swinglehurst et al 2012). Evidence from this research challenges some of the widely held assumptions about the role of interpersonal effectiveness and raises questions about the intense focus on IPS in the CSA when the EPR has already begun to erode some of these interpersonal skills in everyday practice, e.g. the extent of eye contact in the consultation. Other aspects of the CSA, such as time-management skills are likely to be increasingly different from real life consultations with the EPR.

Similarly, globalisation has led to increasing diversity both of the health service and the patient population that it serves (with one third of doctors trained overseas and 29% of the UK population having parents/grandparents born abroad or were born abroad themselves, ref 2011 census) (see 7.3 below). Technologisation, globalisation and the focus on 'soft skills' raises questions about what 'care' means and looks and sounds like. There are now different complex voices in the consultation, i.e. the EPR, interpreters, family members, and different assumptions brought in and brought about as a result of superdiverse patient populations (Swinglehurst et al 2014).

All these factors have led to more standardisation: technology facilitates it and standardisation is widely acknowledged as a means to being seen to be fair and equal in globalised contexts. The value given to interpersonal skills also leads to a quest for agreed ways of defining and measuring them (see Chapter 6). Setting standards in this complex environment is highly problematic, especially in assessing talk and interaction which cannot be codified and standardised on the analogy of standard grammar (Milroy and Milroy 1992) or as scripts (see above). The difficulty of assessing the 'warm skills' of patient-centredness and the interpersonal without making them formulaic has been discussed in Chapters 4, 5 and 6 and, in particular, the paradox that warm skills need to be spontaneous (not trained) if they are to be considered warm. The impact of technology is not addressed in the CSA and diversity is addressed in case design but not in the quality of doctor-patient communication (see below). While these are assessed in the work based assessment component of the licensing exam, it is widely acknowledged that the CSA represents a significant hurdle in the process and has the highest stakes in terms of failure rates. So, both its focus and omissions need to be evaluated in terms of overall fairness to all groups.

7.3   Features of the exam that make it challenging for all

Our research has highlighted some design features which make the exam interactionally challenging for all, although this is by no means an exhaustive list. Some of these features are necessary components of an integrated clinical skills assessment that is testing at a very high level of competency ie that of 'safe, independent practice'.

7.3.1.1   Underlying themes

- a premium on detective work (presenting symptoms may only be the tip of the iceberg) – and the 'what's on the tin' phenomenon. This leads to the assumption that there is always something extra and more complicated to look for which affects the clinical management of the case (and which mirrors many real life consultations);

- balancing professional and institutional elements in relation to the exam, i.e. displaying knowledge, options etc. to the examiner but remaining aligned to patient concerns, expectations (and see 'dual consciousness' below);

- balancing professional and institutional elements in relation to clinical management i.e. the tension between solving a medical problem in the best way and acting as a gatekeeper of scarce resources (e.g. access to consultants, to specialist treatment and drugs);

- dealing with cases where the preferred outcome shows the ability of the candidate to deal with uncertainty, and depends on the justification of the action taken and the number of different outcomes that are possible;

- the time factor in relation to these design factors producing additional pressure, e.g. there may be little time for the right kind of pausing, talk becomes hurried because of this pressure or, conversely, the need to eke out the consultation to last 10 minutes (see Chapters 3 and 5) e.g. recycling and additional safety netting seem to be particularly vulnerable to MAs and MUs.

7.3.1.2   The hidden curriculum and exam modelling

'CSA is a form of regulated communication. There is a strong, if implicit, model.' (Examiner in a feedback session.)

The features just described are all aspects of the hidden curriculum which put a premium on managing the non-medical aspects of the consultation with the RP patient, based on 'common sense' (see Linguistic terms 9) and shared ways of interacting and creating understanding.

The promotion of communication as the key competence, and the close association of soft skills with the cultural self, are part of everyday institutional life. While any consultation is informed to some degree by these discourses, the exam represents a professional ideal which is the product of extensive reflection (Kahneman's system 2 mode of thought which is slow, analytical and effortful Kahneman 2011) and, therefore, like other institutional assessments, reproduces these discourses in a concentrated form as an implicit but foundational model for conducting the CSA.

This model is often unacknowledged or even repudiated but it informs the assessment criteria, the rationalisation for examiner judgements, much CSA training and preparation and is widely used by all candidates (see Chapters 3 and 4) with varying degrees of success. The hyper-noticing environment of such a high stakes encounter has an amplifier effect so that, as the

examiner feedback data show, rapid judgements are made about how candidates talk and relate to actor-patients. The focus on talk and interaction is further amplified by the relatively decontextualised environment of the CSA consultations (see simulation below).

The implicit patient-centred model puts particular focus on IPS (see Chapter 6) – the most subjective, complex and nuanced element of the consultation and the one least susceptible to any standardisation – which leaks into all aspects of the consultation. It is judged in DG and CM domains as well as IPS, so interpersonal effectiveness is assessed more than the other two domains. Even low graded candidates in not being adequately social unwittingly help to refine and reconfirm what a CSA passing candidate should look like and what the institutional norms of the exam consist of. This is illustrated in examiner feedback, e.g. 'he didn't really connect' brings to the surface assumptions about social relations and kinds of knowledge which feed into the specific practices of the exam and into the structuring of medicine as an institution.

So, the exam requires candidates to manage multiple frames: the intersubjective frame of interacting in a social encounter, the modelling of the consultation under the gaze of the examiner and the self-monitoring in relation to assessment criteria.

7.3.1.3   Simulation

In addition to the multiple frames just mentioned, is the additional frame of the simulation – the successful mimicry of a real consultation. This involves more than the realistic content of cases and RPs' portrayal of patients (both of which are rated highly in feedback questionnaires). The candidates (and to a limited extent the examiners) have to work 'at sustaining the 'role-playing' frame as an opportunity for mimicry' (Seale et al 2007:181); for example, the case where candidates and RPs have to improvise about the 'baby' in the waiting room; how candidates have to manage the examiner when they step out of role to give test results; and imagining that things have been discussed before (see the '*MS 1*' case in Chapter 4).

Given the importance of IPS and the emotional response expected from candidates, simulated empathy is a central aspect of candidate performance. Candidates have to experience a 'dual consciousness' in which they have to show enough 'connection' with the patient to be highly marked and at the same time monitor themselves as candidates in an exam (as actors do, Konijin 1997). Arguably, there is an element of 'acting' in all social encounters, a sense we 'perform' our lives (Goffman 1959, Konijin 1997) and there is widespread acceptance that service encounters require skills in managing voice and gesture to perform the 'emotional labour' of professional intimacy (Brown et al 2007). However, in high stakes assessment settings, routine performance has to be raised to a level where it is noticed by examiners and, simultaneously believed as authentic (see Chapter 5). So there has to be a heightened reality and intensifying action to make it visible. This heightened reality, together with a simulated environment, implies that the CSA requires acting skills. Evidence from comments in Pulse suggests that candidates are aware of this and it is also acknowledged by some examiners: e.g. 'I wonder whether he was having trouble acting this particular consultation. He just seemed a little bit kind of remote from him. I just wondered if he was having an acting problem this candidate' (and see Chapter 5).

The hidden agenda element that candidates are expecting can also shift the CSA case from a 'real' consultation to a puzzle or a set up in which both RPs and examiners know what the outcome should be. As a recently successful candidate remarked: 'we don't know which way they are playing it'. For example, RPs are trained to give only a one line opening statement to the candidate whereas in real consultations, most patients give quite an extended presentation of both self and symptoms (Roberts and Sarangi 2004). So the value of a more standardised component to the opening sequence of a case also produces an additional challenge to the candidate (although, arguably, the set format of the opening may be valued for its familiarity).

As suggested above, the simulation is decontextualised in several ways: in relation to lack of continuity of care, the props, networks and information that the surgery affords and, most crucially the absence of a computer screen with which the GP will interact throughout the real consultation. These decontextualised encounters also put specific focus on talk in interaction. The analogy is frequently made with the driving test: 'The CSA is like a driving test – you do things you don't do when driving around routinely.' However this analogy is not quite accurate as the test is taken in real road conditions unlike the simulated cases of the CSA.

Simulation, therefore, adds another frame to the CSA which has to be managed along with the multiple frames mentioned above.

7.3.1.4   Role-player

Given that RPs are not automatons and that candidates are judged on how RPs react to their performance, there has to be some wiggle room. Initial analysis of the data suggested that this gave RPs the power to 'save' or 'sink' candidates, e.g. choosing to surface a possible misunderstanding or letting it pass, e.g. initiating a relevant request after a pause or not. However, as stated above, further analysis supported the research by Foreman (2013) that no one group was disadvantaged by more sink than save examples. There were differences in RP behaviour but overall the training and calibration sustained a high level of consistency throughout. However, simulations do implicate actor-patients in adding to the weight of the exam.

In real life the doctor has the power and patients will do what they can within these asymmetries of power to get their case across. In the CSA the RP has power since their behaviour will affect how the case goes and so how the candidate is judged. (If you know you are likely to fail then you probably feel even more powerless.) In each case, both RP and examiner know how the case should be handled. Interactionally, this power is displayed in the RP interrupting more (as shown in 3.3.2), flagging up misunderstandings more (4.4.1) and was implicit in examiner feedback (Chapter 5).

Role-players are doing what they do well as actors and what is useful for the CSA examiners i.e. acting/reacting in an overt way so that examiners are given more explicit evidence on which to base their judgement than might happen in a real consultation. They may be more helpful or more unhelpful than patients in real life, but because they are actors there will be a tendency

to 'play to the gallery' – not to overact but just make their reactions a bit more explicit than in real life. There was discussion in the examiner feedbacks about the effect of the simulation and how the RPs are reacting. These effects stand out when the process is slowed down and analysed but are much harder to see in real time and where the focus is on consistency within the 13 cases rather than any comparison with real patients.

7.4 Implications for IMGs

It is important to address the fact that focusing on one group, in particular, can itself be counter-productive. Attempting to tackle potential inequality can produce its own stereotypes since it requires a degree of essentialising into demographic groups.

7.4.1 The 'talking' weight of the exam

This weight stems from both the wider discourses that underpin institutional processes (see 7.1) and from the hidden curriculum (the case design and the strong but implicit model of what constitutes patient-centredness) together with the expectation that candidates will talk more (see 3.3). So, for example, many of the tensions outlined above occur in complex cases which require more talk and interaction than more straightforward cases (even when the condition may be severe or complicated).

This focus on communication and soft skills (see Chapter 6, and the frequent priority of linguistic/cultural acceptability over technical skills in face to face assessment generally, see above), and the over-assessment of IPS, concentrates the gaze on just those aspects of IMGs' performance which are hardest for them to acquire and display in the CSA context. What is most intensely looked at (i.e. these communication skills of displaying the self and relating to RPs) is, inevitably, what is most subconsciously processed, as discussed in Chapters 5 and 6. This gaze is intensified by the relatively decontextualised environment of the exam and the tendency for small interactional moments/differences to be amplified.

7.4.2 Overall emotional tone and overall behavioural smoothness

Previous chapters have shown that it is at the micro level of interaction that differences between IMG and UK candidates are most evident. This affects both overall emotional tone – whether the candidate sounds warm, involved, responsive etc – and overall behavioural smoothness – whether the interaction progresses without jarring or uncomfortable moments or not. It is the assessment of the 'manner' of candidates, both how they sound, how they interact and how clear they are in longer stretches of talk that has such an impact on their marks. Sounding different is made up of intonation (and other prosodic features such as pausing, loudness etc.), word stress and pronunciation (often bundled together as 'accent' by non-linguists), as well as other usages, e.g. subtle differences in discourse markers (Hellermann and Vergun 2007:161, and see Chapter 4).

The weight of the exam and International Medical Graduates

These real and perceived differences in the language and interaction of IMG candidates raises questions about the potential for negative evaluation on the basis of language. There is well-established research which shows that in experimental conditions, 'non-native' ways of talking are negatively evaluated – for example non-native speakers are judged as less competent (Lambert et al 1960, Fuertes et al 2012, Gluszek, & Dovidio, 2010) and that stigmatised pronunciation is more noticed if spoken by stigmatised speakers (Lindemann and Subtirelu 2013). While there was no evidence that pronunciation differences were specifically likely to be a factor in lowering grades, other aspects of 'accent' (see below) could unconsciously lead to negative assessments. Recent research in language and social psychology shows that language is a stronger factor than appearance in making negative evaluations (Raki , Steffens & Mummendey (2011) and that implicit evaluations are stronger than any that are spoken out. So there is the potential for discrimination on the basis of language in any high stakes face to face encounter. The findings from this social psychological research could be usefully fed into examiner training. However, there is also recent evidence that suggests that under certain conditions 'native speakers' of English can become more tolerant of perceived foreignness (Hansen, Raki , and Steffens 2014). This latter research accords with the self-awareness of many examiners during the video feedback sessions (see Chapter 5).

Intonation and other aspects of prosody played a crucial role in evaluating candidates, particularly in how formulaic they sounded (see 4.3), in MUs and MAs, how they managed to repair and recover from them (see 4.2) and in more wide ranging assessments of interpersonal skills and clinical management (see Chapter 5). This aspect of talk is notoriously difficult to learn when shifting to a second language or a new variety of an expert language, and requires a long period of socialisation. But our findings also suggest that the particular conditions of the exam put an additional burden on IMG candidates.

The exam model requires explicit patient-centredness and affiliation which has to be managed in order to sound non-formulaic and 'caring' – what we call the overall emotional tone (Erickson and Shultz 1982). This puts particular emphasis on subtle ways of designing and managing how you sound which are readily done by UK graduates but are much harder for those who have learnt English abroad.

More speculatively, the burden of perceived foreignness experienced by IMG candidates may affect aspects of alignment (Chapter 4, section 4.3) and overall behavioural smoothness. This is the ability to manage the interaction fluently and show understanding through sharing words and feelings with the RP; for example, what we have called 'conversationalising', e.g. incorporating patients' words into candidate responses, imagining and inferencing patients' feelings and concerns and so claiming some familiarity with the RP patients; locating and customising empathy tokens and shared patterns of intonation and rhythm. IMG candidates with a different communicative style from the UK trained graduates may come across as intrusive or formulaic when trying to show understanding when quoting back patients' words or inferring and articulating patients' feelings or it is possible that awareness of sounding different may inhibit candidates from using such strategies.

In addition to differences in manner which often affect the emotional tone of the consultation, extended explanations and plans present a particular challenge to IMG candidates (see 4.4). These are central to clinical management and will be a key element in the training materials. Our data show that weaker candidates are more likely to be interrupted during these explanations and that there are slightly more MUs and MAs with IMG candidates. Both of these factors affect the overall behavioural smoothness of the consultation and may feed subliminally into the marks given.

Talk and interactional differences have the potential for being judged as broad cultural differences (see 4.4.2.4). For example, inappropriately designed suggestions to an overweight teenage girl that she might do gym at school, may arise from a lack of cultural knowledge. However, there may be other more immediate social explanations: for example, that it is a communicative issue of hearing and picking up subtle cues or is an exam-induced explanation; that time pressure is hurrying the candidate on and direct, unmitigated suggestions are quicker to ask than more highly designed, 'sensitive' ones. All these explanations are speculative but if broad cultural explanations are taken for granted, these can amplify the sense of 'otherness' for the IMG group.

7.4.3   Simulation and role-players

While UK graduates will have had considerable exposure to simulated consultations, the majority of IMGs will not (references to using simulation in developing consultation skills are recent and rely on UK and North American developments (Pandya 2011).). Before starting training in the UK as GPs, their only experience may have been in the PLAB test. This means that all the factors described in 7.2.3 will present more challenges for them than for UK graduates. The cognitive load required to manage the different frames of the simulation is on top of the socio-cognitive demands of operating in a new variety of English.

The RP ethnic spread broadly mirrors that of the UK population. White majority and BME RPs, whether acting as someone from a different social class or not, in our data shared a broadly similar way of conveying information and attitude and of interpreting others' talk – in sum, the same communicative style, what examiners called 'a BBC radio 4 style', but with some variety in pronunciation (see below for further discussion). Combined with the relative power of RPs (more interruptions, more apologies from candidates) and the magnifying effect of interactional difficulties, the differences in style between RP and IMG candidate contribute to some of their perceived difficulties.

7.4.4   Ethnic and linguistic diversity

IMG candidates regularly consult with patient groups from a similar background to their own. However the CSA does not offer an opportunity to assess the consulting skills and the language and cultural knowledge that IMGs bring to the profession from their earlier background and current practice in the UK. These include consulting in a shared common language and many aspects of linguistic and cultural relativity; for example, tuning the encounter to the

expectations of the patient (such as the extent to which the patient orients to shared-decision making); avoiding British humour, colloquial expressions and metaphors (Roberts et al 2005) appreciating the inappropriateness of explicitly discussing mental illness, knowing that patients' refusal to accept a particular service 'as against their religion' was a product of general fear or ignorance and not sanctioned by their religion (e.g. Kaur-Bola and Randhawa's research (2012) on attitudes to respite care).

While there are BME RPs and tagged diversity cases in the CSA, there is no opportunity for candidates to be assessed consulting in a language other than English or to show understanding of subtle cultural/language differences such as those just listed above (and see 7.4 below). So, IMGs' multilingual skills and communicative flexibility are not assessed in the CSA despite the important social and economic benefits they bring to the practice of medicine, as the widespread use of health professionals as informal interpreters attests.

7.4.5   The IMG experience of the exam

This research has not collected specific evidence from IMG candidates and former candidates and therefore this section can only use evidence from other information to speculate on the experience of the exam for this group. However, since exam performance depends so crucially on this experience – both what candidates bring to the exam and how they experience its weight – it must be taken into account in any study of what leads to such differential outcomes for the different candidate groups. Research has shown that when a stigmatised racial or ethnic group has knowledge of the possible stereotypes associated with their performance that this will affect their performance (Epley p. 137 2014, quoting Steele and Aaronson 1995).

This section draws on examiner feedbacks, the Luton video (see Kanchandani (2011) and appendix D-3), comments in Pulse, Mohanna's (2011) Ed D preliminary study and several conversations with IMG candidates and GP trainers at conferences and in more informal settings.

Both the knowledge that IMG candidates are far more likely to fail than UK Graduates (15 times more likely) and the personal experience of failure has, according to the sources given above, led to anxiety, anger and fear (see Chapter 5 on the 'terrifying experience' of the exam as perceived by some examiners).

The research and published comments just mentioned indicate that many registrars perceive the CSA as requiring acting skills, prioritising language and inter-personal skills over other skills and not taking account of IMGs' capacity to care for patients from ethnic and linguistic minorities. IMGs may not appreciate the part that communication and interpersonal skills play in the eyes of the examiners and trainers or may think it is over-emphasised to discriminate against them. There is a similar phenomenon in other English speaking countries such as Canada, Australia and New Zealand where professionals cannot see why their expertise is not valued and see the soft skills demanded as setting too high a hurdle for them to jump over (Allan 2013).

There were also comments that the exam genre propels candidates into formulaicness: 'If we adhere to a structure, we can't help being formulaic. Adhering to a structure helps us but (we)

have to try not to be formulaic' (Luton video). Registrars and trainers talked of an atmosphere of 'doom', of feeling demoralised, lacking in confidence, often exacerbated by preparation courses where candidates may be encouraged to use set phrases.

There were differences of views about how IMGs thought they could be prepared for the CSA. Some registrars still turned to studying more, but the majority of registrars and trainers were adamant that the practical knowledge and skills learnt from consulting day to day in the surgery and reflecting on these consultations was the only effective preparation and not studying more: 'Having too many books about the CSA – repetitive. Just don't go to theory all the time. I failed and studied more and more' (Luton video Example D-3). And both registrars and trainers mentioned the importance of practice and reflection with mixed groups of local and IMG registrars. This suggests that social capital (the networks, norms and trust that enable participants to act effectively together to pursue shared objectives), should not run on ethnic lines but be developed as part of all registrars' mutual support.

7.4.6   Ontological and linguistic insecurity

The sense of doom and demoralisation can affect the confidence and self-belief of IMGs when what trainers identify is the need for more openness and reflexivity, qualities which themselves stem from self-confidence. The anxieties and negative experiences of the exam can feed into more profound feelings of insecurity. Mobile professionals, such as IMGs with English as their expert language in their country of origin, experience disjunctures between their habitus (their ways of doing and valuing things) and the new fields they encounter. Initially, this new field is the NHS and subsequently the CSA. The resulting disjunctures can generate not only change and transformation, but also disquiet, ambivalence, insecurity and uncertainty (Reay et al 2005 and see the introduction to this research). This is not the result of any narrow definition of 'language difficulties' but rather relates to the subtle communicative requirements described in earlier chapters (Jain and Krieger 2011).

These insecurities, in turn, can have a profound effect on what social theorists call the on-going narrative of the self (Giddens 1991) or the self as a project (Goffman 1959). While there is always some distinction between the continuous self and the performance self (Giddens 1991: 58) (for example, even the highly successful group of white female candidates have to put on something of a performance), the fit between this coherent, continuous sense of self and the performance is crucial to a secure sense of being. Where the weight of the water bears down and the candidate feels a fish out of water, there will be less fit between a sense of self and the performance required and the candidate may feel more the pressure of having to act and, in the longer term, more what Giddens calls 'ontological insecurity'.

Ways of talking and interacting are central to this sense of self. Where there is a gap between habitual ways of talking and how you are told to talk, then self-identity is challenged as it is difficult to integrate this new regulated talk with an on-going narrative of the self. When attempts to address this 'lack of fit' centre on regulating communication, then self-identity can be further challenged.

There is a double challenge for IMG candidates. Firstly, they have to learn the exam genre and its implicit model. Secondly, the modelling and standardisation of the exam have to be counter-balanced by conversationalising and customising the IPS component, whereby meaning and sociality are largely communicated indirectly and informally and grounded in background understanding. The more distant the models and assumptions of the exam are, the more explicitly they have to be called up and used. This leads to what we have called over-modelling. In doing so, there is less indirectness and so less sociality and more sense that the candidates are not drawing on shared backgrounds with patients.

So, although IMGs are expert English speakers in their countries of origin, 'linguistic insecurity' is added to 'ontological insecurity'. IMG candidates have to laminate exam talk onto their own ways of speaking/interacting and try and sound 'caring'. This can produce a hybrid which is judged as lacking authenticity and sincerity, hence the comments of 'formulaic', 'clunky' etc. By contrast the personal 'voice' of candidates trained in the UK coalesces with both the institutional talk of the exam and its more personal and informal element.

The insecurities experienced by IMG candidates both before and in the exam may serve to explain some of the performance factors. For example, there is a frequent criticism that poor IMG candidates do not control the consultation enough, do too much referring, do not take responsibility for making a diagnosis or are over-compliant. This tends to be rationalised as 'culture' and given a cultural explanation that not being too firm and being humble are perceived as admired qualities in Asia. Similarly, there are criticisms that talk is rushed and questioning too fast. In many cases, these performance features may not be the result of broad cultural differences but rather to do with nerves and a more profound sense of insecurity as well as the social factors created by the exam itself (see 6.3.2 above).

7.5 Implications for UK graduates

The weight of the exam bears down much less heavily for most UK graduates. This is for many complex reasons beyond the scope of this research but include: (i) their ways of talking and interacting are compatible with the exam genre (ii) their language/cultural skills with linguistic minority patients are not assessed and (iii) they are experienced at handling simulations.

7.5.1 Ways of talking compatible with the exam genre

As Chapters 4, 5 and 6 show, successful candidates talk to the exam model but can customise and conversationalise interactions so that they sound caring and sincere but still able to control the consultation. For example, they use more vague language which also accords with the findings from the corpus of white doctor/patient real consultations (see Chapter 3). Early socialisation provides linguistic and cultural capital and undergraduate and postgraduate training provide considerable exposure to GP practices. They are fish in water in the CSA, able to draw on readily available linguistic resources to manage the consultations successfully.

7.5.2   What the tagged diversity cases assess

As mentioned above (Section 2.3.5), the CSA case bank cases are tagged for diversity, including social class and disability and also RPs are drawn from black and minority ethnic backgrounds (BME). In the data collected in the Feb-March and May 2011 diets, 9 were tagged for diversity with 'patients' from a BME background, only one of which concerned a patient for whom English was a new language.

A review of these 9 cases did not show any interactional differences between these cases and those not tagged 'diversity'; nor did the cases with BME RPs show any interactional differences when compared with local white RPs. Often the candidate was required to show cultural sensitivity, e.g. offering a female chaperone and, e.g. a case played by a black RP with a condition more commonly found with these ethnic minority groups. In one case, the RP was asked to act as a Polish patient with limited English. However, this case was dropped because of the difficulties RPs had in sustaining the accent and reduced fluency in English throughout a day of repeated cases.

When compared with real consultations with a multilingual patient population (Roberts et al 2005, Roberts and Sarangi 2004), none of the following were evidenced:

i.   Persistent misunderstandings arising from patients' differences in pronunciation, intonation, grammar and vocabulary

ii.   Persistent misunderstandings and information loss because of patient lack of cohesion or clear line of argument in their narratives

iii.   Clear differences in patient presentation of self and symptoms in the opening phase and also difficulties GPs found in closing the consultation with this group

iv.   Different interactional patterns, e.g. GPs' attempts to share information and decision making interrupted by patients introducing new symptoms

v.   A range of GP strategies to prevent and repair misunderstandings.

vi.   Misalignment when humour and other attempts at social chat failed.

The contrast between the CSA diversity cases and the real cases with linguistic minority patients results from the RPs' style of communicating, which is primarily a 'a BBC radio 4' style, as mentioned above, with some more role-played reticent individuals. While, if required by the case, RPs can act 'patients' from different social classes, their jobs as actors mean that they talk and process talk as local English speakers. The lack of success with the RP with limited English is a case in point. The type of interactional differences listed above cannot be simulated with any degree of authenticity.

This review of diversity cases indicates that although candidates should be able to practise anywhere in the UK, they are not assessed on consulting with linguistically diverse patient populations, including working with interpreters. As cities and towns become more diverse, and research on these groups sheds light on the problems they face in GP consultations, the

skills required become more important (e.g. Roberts et al op.cit., and Schouten et al.'s research (2009) which shows that there is less socio-emotional exchange between linguistically and culturally diverse patients and they are less involved in decision making). The CSA bandwidth is not inclusive of many of the cases that GPs will have to manage and from our data, the more complex cases were generally with assertive, informed, demanding, middle-class actor/patients (e.g. the *'MC'* and *'MS'* cases etc.). So, UK graduates are not assessed in an area where their language/cultural skills could be challenged.

7.5.3   Simulations

Candidates trained in the UK are likely to have been exposed to role-play as a type of learning and formative assessment from their school days and will have practised consulting skills with actor-patients from at least the second year of their medical degree. They will also be habituated to the multiple framings of such activities. In OSCEs they will have had experience of simulated consultations in summative assessment.

It is also the case that those who are most experienced in simulation are those for whom the exam is closest to their own practice, sense of self and presentation of self. They can be friendly, charming and (relatively) at ease because they are not having to re-engineer themselves for the exam.

7.5.4   UK BME graduates

There has been considerable research but no definitive explanations as to why BME UK undergraduates and graduates fair less well in medical assessments. Our research looked at cultural/linguistic performance factors across all three groups: white Anglo and BME UK graduates and IMGs. What clearly stood out was the range of subtle talk and interactional features which marked the difference in performance between UK and international graduates. No such obvious differences stood out between white Anglo and BME candidates. However, within our data base, we noticed a small number of candidates who were not categorised as IMG but had some of the characteristics of talk and interaction which were observed in the failing IMG group. Some CSA candidates will have come to the UK to take up an undergraduate place but whose education and early socialisation were in another country. This may be a contributing factor in the somewhat larger number of BME UK graduates who fail as compared with white UK graduates.

Chapter 8.
Educational interventions

8.1   Summary

Educational interventions aimed at registrars and GP trainers, but also useful for examiners and those designing the exam, will be produced. These will consist of a set of e-learning materials: 'Looking Inside the Clinical Skills Assessment'. These materials will be based on the following principles: (i) the use of real video recorded CSA cases (ii) the introduction of a new analytic language (iii) awareness and self-monitoring developed from close video analysis (iv) awareness raising and tolerance of different varieties of language use (v) identification of cases in terms of interactional rather than medical difficulty. These materials will also be linked to a book on the CSA by Alex Rolf and the KTP team to be published by the RCGP.

Developing skills over time requires constant practice, in an organised and supported way and involves embodied experience, relational understanding and a sense of incompleteness in a positive self-critical way: 'Skill is a trained practice ...Going over an action again and again enables self-criticism' (Richard Sennett 2008:38 -9). These three elements: practice, support and a self-critical understanding provide a useful framework for educational interventions for trainers, candidates and examiners. This chapter outlines the development of planned materials that focus on the interactional and language demands of the CSA: 'Looking inside the Clinical Skills Assessment' and also discusses what aspects of communication cannot be readily turned into training materials.

Previous s have shown how subtle, complex and dynamic institutional interactions can be and that much meaning making goes on subliminally and is easily overlooked. These are elements of the CSA that are part of a hidden curriculum consisting of tacit knowledge and skills. This creates two problems when addressing training, preparation and examining. Firstly, many of these unanalysed elements are difficult, perhaps impossible, to make explicit and learn. Secondly, as the problem with judgements of being formulaic have shown, there is a paradox. By bringing these more hidden aspects to the surface and making them amenable to practice, they sound 'trained', jarring and insincere. Regulating communication can also lead to insecurities about the self (see Chapter 7). For this reason, we outline some general principles underlying the materials which we hope justify our approach, while acknowledging the difficulties of tackling this area.

Educational interventions

8.2 'Looking inside the Clinical Skills Assessment': some general principles

8.2.1 The use of real (and not simulated) materials

In line with our discussion in Chapters 5, 6 and 7 on some of the difficulties of using simulation in summative assessment, the use of real rather than simulated materials in formative assessment is proposed. In this case it is the real experience of managing a simulated consultation which is used.

10 successful candidates (from the 40 focal cases) have consented to the use of segments of their video recorded consultation. Some of the messiness of interaction, the nerves caused by the exam and its time pressures are visible and can be used to discuss what seems to count in making assessments.

8.2.2 Analytic framework and language

Ethnographic evidence from discussions, feedbacks, case guidelines and comments has shown that there is no shared and agreed language for talking about the non-clinical aspects of the CSA. This affects how GP trainers and candidates manage the training and preparation both for the CSA and for consulting more generally. It also affects how examiners make assessments, give feedback and are trained. There has been concern that the feedback statements are too general to give specific guidance on what needs changing and the RCGP is already working to improve this area (and see O'Grady 2011 for a similar finding in Australia).

This research has drawn on a series of terms from sociolinguistics and discourse analysis which give some analytic building blocks and draws attention to some of the details of talk and interaction which are overlooked. They include:

- Alignment and misalignment (rather than empathy, 'clunky', connection)

- Narrative, metaphor and cohesion – to analyse the communicative aspects of explanation and other extended stretches of talk

- Inferencing, repair and metacommunication (to explain aspects of interaction – listening and responding, rather than the general term 'active listening')

- Prosody – an understanding of intonation, pitch and tone (some of the music of language which conveys both attitude and information and is used differently in different varieties of English).

8.2.3 Detailed awareness through close video analysis

The mundane aspects of communication are not readily noticeable. Close video analysis puts the microscope on talk to reveal these aspects. So a habit of slowing down and looking at the detail is as important as using a new analytic language. This slowness and smallness is effortful

and requires the analytic language just mentioned. It also helps to reveal that what is being judged is a joint activity – that there is a constant interplay between conversational partners which affects what the other does. Training is often centred on what registrars should say and do rather than seeing the consultation as interactive (Frankel 2009, O'Grady 2011).

8.2.4 Aspects of communication not susceptible to explicit teaching

Teaching set phrases or attempting to change how individuals sound can lead to the equivalent of painting by numbers. As many trainers and registrars recognise, there are aspects of the consultation that require a gradual process of socialisation rather than explicit training. This research has shown that prosody is powerful in: conveying a particular attitude (magnified in the patient-centred model), managing self-corrections, coherence in explanations and in signalling shifts in speaker purpose. This aspect of talk is culturally-specific and different varieties of English have different norms in interpreting how prosody is used. It is helpful for all three groups to understand this aspect and to recognise that it cannot be readily taught. However, attending to these mundane but important features as part of a larger process of socialisation into interactional norms is useful. And for examiners and CSA designers, this knowledge can shape decisions and judgements about the exam genre.

8.3 "Looking inside the Clinical Skills Assessment": Specific training areas

8.3.1 Key aspects of the CSA

- interactional and affiliative alignment

- customising the patient-centred model

- avoiding and repairing misunderstandings and misalignments

- managing personal and institutional stance from the outset

- aspects of giving a good explanation

8.3.2 Using a new analytic language

Chapter 4 gives examples of how to do close video analysis. They include the use of new analytic terms and developing a habit of smallness and slowness when looking at either your own or other's video recorded consultations or the CSA materials in the training modules.

This analytic language and habit of looking closely and slowly is important for all registrars. As well as helping IMG registrars, this habit is important in developing skills and awareness in UK graduates dealing with patients with different communicative styles from their own. (For an example of real GP-patient consultation video segments in a superdiverse setting, see Roberts

et al 2007). This new analytic language may be difficult to understand, at first, and the materials will make a special point of illustrating these terms in a variety of ways so that they become familiar.

All three groups (registrars, trainers and examiners) may also find it useful to compare the close analysis of CSA segments with real consultations to highlight aspects of the exam genre different from everyday consulting. In concentrating on communication skills, it is also important to recognise that in the day to day life of a GP practice, interactions may be more convivial and both sides more tolerant of each other. Outside the standardised context, in our day to day jobs there is more scope for diversity in interactional styles and we do not always get our best or most difficult work done through an interactional 'ideal'.

8.3.3 Analysing different types of cases

The materials will develop an awareness of different types of cases, their structuring and the different stances that candidates need to take. Categorising cases in terms of communicative complexity and sensitivity and not just in terms of medical complexity is important for preparing candidates. Complex cases are often agonistic (see Linguistic terms 2), i.e. require an argument or persuasive element which affects the patient-centred stance and may require a more institutional stance.

Chapter 9.
Conclusions and implications

9.1 General Conclusions

- This research addresses a central paradox in institutional life: how to be fair to a diverse group of candidates in an increasingly diverse society, while maintaining standards and universal criteria. This is a profound and difficult area to address since assessment of consulting skills is through largely subjective judgements of face to face encounters which produce a flow of reactions and impressions.

- The CSA consists very substantially of talk and interaction. Sociolinguistic analysis provides a rigorous methodology for evaluating performance and contributing to an understanding of how to close the gap between the different views on how fair it is. Looking at the 'how' of performance and not just its outcomes can also steer consulting skills training and preparation for the CSA towards a more linguistically informed approach.

- There are limits to sociolinguistic analysis and the ethnographic methods which support it. Using a mixed method approach, largely qualitiative but with quantitative aspects can make research vulnerable to criticisms from both positivistic and interpretive traditions. Our goal was to be as holistic as possible and avoid decontextualised abstractions, on the one hand, and over-specificity on the other, while producing a series of telling cases with broad implications. So there are limitations that relate to the size of the data base for corpus linguistic analysis and coding and the small number of examiner video feedback sessions. There are also limitations to the micro-analysis since the 40 cases represented a large data base for such analysis and so only certain features could be identified.

- The research shows that (apart from strictly medical aspects of the exam) the differences between successful and unsuccessful performances are only clear at the micro level of analysis, with two exceptions related to late data gathering and explanations.

- At the micro level, the clearest weaknesses were in extended explanation and other long stretches of talk, managing the patient-centred model of the exam and managing the somewhat more frequent misunderstandings and misalignments. These areas presented a greater challenge to International Medical Graduates than to UK graduates and the IMG group need more support and guidance in these areas.

- The CSA is a culturally specific exam. The non-medical aspects are judged on largely mono-cultural norms of talking and interacting and, although there are 'diversity' tagged cases, the exam does not reflect the diversity of contemporary British society in terms of interactional styles or how patients present themselves. So while the rationale of the CSA is that it will assess candidates to ensure they are fit to practise anywhere in the UK, at the moment, the exam does not assess how well all candidates can practise in multi-lingual urban sites. These factors need to be considered in reviewing the CSA.

Conclusions and implications

- These norms are most powerful in the interpersonal skills domain (IPS) which is evaluated as a domain in its own right and across the other two domains (data gathering and clinical management) and is also the area most linguistically demanding for IMG candidates. The IPS domain presents particular problems and our evidence shows that 'empathy' and 'rapport', in particular, may not be best assessed in a highly standardised exam of simulated consultations.

- IMG candidates are also more likely to experience the weight of the exam because of the unintentional, or hidden, effects of the factors mentioned above and also the additional communicative demands of coping with simulation and the exam genre. Indirect and unintentional processes can lead to perceived and real inequalities and the purpose of this sociolinguistic analysis was to investigate the processes of some of these.

- However, there was no clear evidence that RP behaviour systematically led to any one group of candidates being 'saved' or 'sinking'.

- Both the difficulty in meeting the expected demands of the exam and the unintentional weight of the exam, form some of the moving parts which contribute to an understanding of why (apart from clinical omissions and errors) IMG candidates fare so much less well than UK graduates. Both aspects need to be addressed through (i) educational interventions (ii) consideration to changes in the exam design and how consulting skills are assessed.

- Educational interventions aimed at candidates, GP trainers and those involved with designing and examining the CSA, and based on the research findings, should provide a consistent analytic approach across all these groups based on sociolinguistic methods. This should ensure that candidates know in detail the communicative challenges of the exam and have specific guidance on how to develop the necessary resources.

- The exam amplifies differences between UK and IMG candidates because of the intense focus on aspects which IMGs may find most challenging such as the norms of interpersonal effectiveness and not assessing aspects that UK graduates, particularly white British candidates, are likely to find challenging.

- Rather than talk of cultural bias or not, there needs to be a debate about tolerances and communicative flexibility, about what limits should be put on them and how within these limits, talk and interaction can be more explicitly addressed

9.2 Anatomy of the non-clinical performance features

9.2.1 Comparison of successful and unsuccessful candidates

These conclusions are based on mapping the 40 focal cases.

- There is little difference in the overall structuring and timing of cases between successful and unsuccessful candidates. However late data gathering was marked down as 'disorganised/unstructured', even when its purpose is to repair earlier omissions. This raises questions about whether mistakes or missteps should be judged more tolerantly if they are subsequently repaired.

- Successful candidates in routine consultations talk on average 68% of the floor time. This is a little more than in 'everyday' GP consultations and shows that this is a 'talking heavy' exam.

- While there are no great differences in levels of interrupting between pass or failing candidates, interruptions in the explanation phase is a clear marker of less successful candidates. These function to highlight inadequate explanations and also disrupt the speaker's flow and speech planning, so making the explanation hard to give. So explanations are a particularly demanding aspect of the consultation.

- There are some words and phrases that are particular to the CSA, suggesting strong 'formulaic' differences from aspects of everyday spoken English, and some difference from everyday GP consultations. These show a clear CSA 'fingerprint'.

- These formulaic expressions cluster around the social/interpersonal work of the CSA and show that it has a strong patient-centred model.

- Successful candidates tended to use more exam-modelled strategies. However these were just the strategies that were also deemed 'formulaic' in examiner feedback. (This paradox is explained below).

- Role-payer talk reflects the highly 'conversationalised' English style of both patients and health professionals in institutional talk. However, it cannot reflect the variety of Englishes and other languages heard amongst multilingual patient populations (see Linguistic terms 8). This may disadvantage candidates who regularly consult with such a population and advantage others who do not.

- Successful candidates showed marginally more alignments and marginally fewer misalignments than unsuccessful candidates.

- Overall, the lack of difference between candidates at this general level suggests that significant judgements are made at a much more micro-level to build an accumulative impression of candidates' talk.

- Small interactional differences can have large consequences both in how the consultation plays out and in the judgements of the examiners.

Conclusions and implications

9.2.2   Successful features of the CSA at a micro level

- The term 'alignment' is more useful than 'empathy' or 'rapport' in developing a shared language that is based on observed behaviour and is not about making assumptions about people's inner feelings. While there is considerable 'affiliative alignment' i.e. explicit strategies of caring and agreement rated as 'patient-centred', even more important is 'interactional alignment' – the accumulation of small moments which make the interaction comfortable and relatively agreeable. Alignment is also a joint production between candidates and role-players and the latter's responses are a powerful channel through which a candidate comes to be judged.

- Alignment is happening throughout the consultation. The following were identified as important in interactional alignment in all three domains: accurate listening and inferencing of small and subtle cues and the use of conversationalising strategies i.e. small markers that make the consultation more informal and more sensitive to the listeners' stance in order to provide the consultations' 'social glue'. These include: making language more indirect, (e.g. using 'softeners' such as 'could', 'might') more light-hearted and more vague.

- Affiliative alignment strategies function to show the candidate is patient-centred. These relate primarily to establishing patients' issues, concerns and expectations (ICE), to their psychosocial world and to checking patient understanding. These 'empathy' tokens correlated with success but only where they were located appropriately in the turn by turn interaction, customised to avoid sounding formulaic and used expressive non-verbal features.

- Successful candidates also used key features of a good explanation: signalling logical connections in definitions, using repetition to list causes, symptoms etc., making explanations cohesive, using metaphor, narrative, metacommunication and relating back to patients' concerns and language.

- Other factors which affect the degree of alignment relate to the particular features of the exam, i.e. the relative power of the role-player, the patient-centred conventions of the exam and the need to rapidly imagine aspects of the case which have not been given in the notes to the candidate. So 'alignment' depends as much on managing the exam genre as simulating a comfortable encounter.

9.2.3   Unsuccessful features of the CSA at micro level

- Virtually all candidates experienced moments of misunderstanding (MU) and/or misalignment (MA). But IMG candidates and poorer performing candidates had slightly more (see 3.5 above).

- Repair is therefore a very important feature of the CSA since some form of communicative trouble is frequent and attempting to repair it is more noticed by examiners. Where the trouble was repaired swiftly, located appropriately and did not

sound as if it stood out as a problem, it could be rapidly defused and candidates did not appear to lose marks in the IPS and clinical management domains because of it. In all other cases, it was difficult to recover from mistakes or missed opportunities. Intonation as well as location could be a factor in IMG repair (see below). If repair is successful, even if it led to awkward moments, the successful negotiation should be valued and used in mitigation.

- MUs were caused by mishearing and forgetting, missing clinical information and exam – modelled sequences where RPs' responses and questions led to off the cuff explanations and re-runs which themselves could cause more MUs. These were not caused by pronunciation differences (generally described as 'accent') but intonation and other prosodic features led to some MUs among IMGs (see below).

- MAs resulted from MUs but were also caused by disfluencies in speech and causing patient alarm. Disfluencies occur in all talk but are likely to be caused by the exam context and this raises the question of whether a more tolerant approach should be adopted.

- Role-players flag up misunderstandings in the CSA more than patients tend to in real life GP encounters and candidates tend to be on the back foot apologising. But this is a product of their role in a high stakes assessment, and we found no evidence that they were behaving in a systematically unfair way.

- The line between achieving alignment and misalignment is quite a fine one. Many candidates (successful and unsuccessful) use the same/similar strategies – but with differing levels of success. In particular, the subsequent responses of the role-player seem to mark it out as successful or not. Many of the affiliative alignment strategies used can backfire and cause misalignment – and some open ended questions actually offer the opportunity for misunderstanding, which seems to have a negative effect in this simulated setting. Again location and precisely how questions are posed by candidates or attempts at repair made, often make the difference between marks being lost or not.

- So although most candidates use the same patient-centred questions and 'empathy' tokens, unsuccessful candidates are routinely judged as being formulaic and this contributes to lower IPS grades. It is particularly challenging for candidates whose English is influenced by other languages not to sound formulaic. Since the largest group of IMG candidates is from the Indian sub-continent, this research concentrated on certain features of Indian English which has been quite widely studied. Intonation and other prosodic features may be at least partly used differently in Indian English to express affect (see Chapter 4).

- Similar features of Indian English and other Englishes influenced by a first language can lead to communicative differences which in extended explanations can cause difficulties in the weakest candidates in tracing the line of argument, processing how

phrases and clauses link together and managing a conversational tone. All candidates display some of the features of successful explanations but the lack of local coherence and explicit tying of ideas together can undermine these successful strategies.

- Our micro analysis shows that speakers and listeners are tuned in to small features of language, and, beyond words, intonation and other ways of delivering talk which are fleetingly processed but contribute significantly to how the interaction proceeds, how the role-player patient responds to the candidate and how the candidate is judged by the examiner. Such features are present in all languages but may have quite different uses and values across different languages.

9.2.4   Interpersonal skills and subjectivity

- Setting aside strictly clinical diagnosis and reasoning and decision-making, all other aspects of the CSA are judged subjectively as with all face to face assessments. This is an inevitable and entirely understandable process and should not be tarnished by the term 'subjective bias'. People judge each other by how they come across, their manner, and so their competence and attitude. Some aspects of candidate behaviour appear considerably more noticeable than others, e.g. certain types of bodily conduct and unexpected or different ways of talking, in particular how candidates sounded. In the examiner feedbacks, these were frequently extrapolated out from the interaction and led to broad evaluations.

- 'Bias' implies that there is a goal of neutrality that must be reached and a dichotomy of biased/unbiased. Rather, in primarily subjective assessments such as the CSA, it is more useful to consider performance in terms of tolerances – the reasonable leeway for variety, and context – the range of experiences in daily GP practice – and the relationship between tolerances and contexts.

- There was considerable agreement in rating candidates' performance in the areas that examiners chose to comment on i.e. largely about manner and affect. These shared interpretations of candidate behaviour are based on inter-group subjectivity, drawing on 'common sense' and professional shared meanings of this relatively homogeneous group.

- However, there were also individual differences which are likely to be based on individual style and individual noticing of aspects of behaviour and are part of the inherent subjective processes of evaluating interaction. These individual differences are likely to be ironed out across the 13 different cases but the inter-group subjectivity remains.

- While subjective assessments feed into all three domains, since all are accomplished through talk and interaction, the IPS is specifically designed to look at manner and inter-subjectivity and is therefore the domain least likely to be leavened by standardised knowledge and procedures.

- While the IPS domain reflects current assumptions about patient-centredness and a warmer, more equal relationship with patients, this research contributes to the on-going concerns expressed about how to assess and measure it. IPS is assessed in terms of a patient-centred model and the overall manner of the candidate. The problem is that in order to be assessed, interpersonal effectiveness has to be displayed through talk such as 'empathy' phrases. But this explicit talk can produce the opposite effect since a shared emotional tone is substantially created through more indirect means such as humour, metaphor and an informal conversational mood based on tone of voice.

- The implicit weight given to models of patient-centredness and IPS may not be appropriate for all patient groups. In particular, patients from linguistically diverse communities may have different expectations from the majority patient group about the doctor's stance and ways of relating. The focus on patient-centredness may also deflect candidates from noticing and acting on managing the tension between patient-centredness and the institutional requirements of some of the CSA cases.

- Manner is assessed through small but cumulatively significant differences in how meanings are conveyed. These are the most culturally specific aspects of language and attempts to repair misunderstandings may exacerbate the difficulties. Examiner criticisms of formulaic and 'clunky' performances arise from this focus on manner.

- The subjective and culturally specific nature of IPS assessment affects candidates twice over. So the IPS domain is assessed more than the other two. This is evident in (i) the marking statements for the cases where IPS overlaps clinical management and data gathering much more than either CM or DG overlap with each other or IPS (ii) judgements of clinical management being regularly conflated with assessment of manner and affect in examiner comments and (iii) the microanalysis of interaction which shows that the basic conversational/interpersonal skills of most interaction feed into the specific skills of diagnosing, displaying, clinical knowledge and clinical decision making.

- The exam itself produces occasions for the interaction to appear uncomfortable or jarring. Textbooks and CSA preparation and training courses can feed into assumptions that it is a case of rehearsing appropriate phrases to show this model. Managing the exam model is not easy, and while all candidates display knowledge of this model, it requires additional communicative skills to massage it into an easy conversational flow, rather than produce a learnt phrase. These additional communicative resources (ie displaying empathy through explicit phrases but then customising them) are necessary for the exam but not necessary for real life practice.

- These findings raise questions about the weight given to IPS and the standards by which this domain is judged. On the one hand, aspects of IPS are assumed to be readily and explicitly assessable when our research suggests they are not (eg 'empathy' and 'rapport') and, on the other, all domains are implicitly assessed through an IPS frame but are not recorded as such. The fact that very little or no guidance in this area is given

in the case descriptions suggests that this is a taken for granted area of competence for examiners.

- A hypothetical test run to assess the effect of removing the IPS domain from the assessment showed that the marks of both IMG and British candidates would be affected but only marginally i.e. 3% more of IMGs would pass (4% of male IMGs) and 2% more British candidates would fail. However, this did not account for the leakage of IPS into the other two domains, particularly clinical management.

9.2.5   Aspects most challenging to IMG candidates

The higher failure rate of IMG candidates can be attributed both to certain features of their performance and to the exam-induced effects on their performance.

Performance

- This research excluded strictly clinical medical aspects of CSA performance and currently there is no research which systematically identifies the extent to which candidates fail primarily for these reasons (although written examiner feedback statements give some indication of this). Recent RCGP analysis has shown that it is in the area of clinical management (CM) that failing candidates perform least well. However, our evidence shows that there is considerable overlap between IPS and CM and that since manner and affect are assessed in all three domains, CM is not narrowly concerned with displaying clinical knowledge and reasoning. So IPS, the area most susceptible to linguistic/cultural judgements, is assessed more than the other two domains.

- As detailed in Chapters 3 and 4, successful candidates display a range of strategies that include: managing the patient-centred model, conversationalising, clear and patient-involving explanations and managing repair. Lower performing candidates handle these aspects less well, e.g. are more likely to be interrupted when giving explanations.

- While all candidates may show some weaknesses in these areas, they present particular challenges to IMG candidate; for example, this group had slightly more misunderstandings and misalignments, whether they were passing or failing, than other candidates.

- Candidates' 'manner', i.e. how they sound, manage longer stretches of talk and deal with interactional problems are central to the assessment. So those who are using a variety of English that is different from the great majority of candidates are vulnerable to more negative evaluation. While there was no evidence that pronunciation as such was a factor in these ratings, intonation and other aspects of language delivery did feed into negative evaluations (see Chapter 5). It is not helpful to talk about language deficit or needs in a general way. This does not respect the expert use of IMG candidates' English. A much more nuanced understanding of language varieties and how they are judged is necessary.

The weight of the exam

- The exam is set within some of the wider themes and assumptions of today's society: a focus on communication and the interpersonal, increasing use of computerised records and increasing cultural/linguistic diversity. While the first of these plays a dominant role in the CSA, the latter two remain marginal to the exam.

- IMG candidates feel the weight of the exam in ways not experienced by UK graduates. This stems from the focus on manner and the interpersonal, the decontextualised environment, its simulated nature and the fact that the great majority of examiners, all RPs and UK graduates have grown up sharing a broadly similar use of English, thus giving a homogeneous feel to the exam. These factors make up the 'hidden curriculum' of the exam.

- Many features of the 'hidden curriculum' of the CSA put a premium on managing the interaction intersubjectively (see Linguistic terms 9) and so, on interpersonal effectiveness. This concentrates the gaze on just those aspects of IMG performance which present the greatest challenges to this group.

- In addition, the decontextualised nature of the consultations, in particular the lack of doctor-computer interaction and the limited opportunities for physical examinations, also focuses the gaze on talk and interaction, evidenced by the fact that candidates talk more in the CSA than in real life consultations.

- Simulation presents particular problems to IMG candidates, in particular: the differences in communicative style between most IMGs and RPs; the lack of experience of role-playing at school or in their undergraduate education; the lack of opportunity to assess IMGs' regular experience of consulting with BME groups with whom they share a cultural/linguistic background.

- Simulation adds to the complexity of the exam, requiring candidates to manage multiple frames: modelling patient-centredness, reacting in a real and intense social encounter and managing the simulation. This requires both mimicry and a heightened reality. Managing multiple frames puts an additional burden on candidates' communicative power. Simulations should be used for what they do well, not what they do less well. Measuring standardised interpersonal skills in high-stakes exams is not one of them (see Chapter 6).

- Most UK graduates possess the linguistic/cultural capital required of the exam. They are also advantaged in that: the diversity cases in the CSA do not assess their cultural/linguistic skills in consultations where they might be 'fish out of water' with patients from linguistic minority backgrounds; they also have considerable experience of simulations and so a feel for the game.

- So, several aspects of the exam amplify the differences between UK and International Medical Graduates. Many aspects of consulting in the real world are converted into highly valued aspects but others are not. Similarly, some elements of the exam play

too great a part. The interactional and interpersonal are filtered through a rather confined notion of what is acceptable, whereas examiners suggested that linguistic/cultural differences are likely be treated in a more convivial and tolerant way in real life situations.

- There is no evidence that RPs consistently contribute to 'sinking/saving' particular groups of candidates. However, they do have power in the exam and very small interactional differences can affect candidate performance.

- A small number of UK BME graduates in our sample displayed some of the differences in 'manner' of many of the IMG candidates. They may have been schooled abroad and only educated in the UK from undergraduate level (although there are no records to dis/confirm this). This might contribute to the failure of a small number of UK BME graduates.

What IMG candidates bring to the exam

- Evidence from other research (see Chapter 7) suggests that IMGs experience anxiety, anger and fear because of their high failure rate and this is likely to lead to insecurity about themselves and about their competence in communicating which affects their performance.

- IMG candidates bring considerable experience of consulting with linguistic minority patients with whom they share a common language. They also bring the experience of having to change their styles of consulting, of having to re-engineer the variety of English that has made them authoritative English speaking professionals in their own country, of having to deal with cases which are different from those they are used to, of having to manage a style of exam which they are unused to. They also bring the knowledge that they have a good chance of failing and may, like professionals in other countries, be critical of the emphasis on soft skills at the apparent expense of bespoke professional skills. All these experiences and the values derived from them are likely to affect their performance and may well account for some of the disfluencies and uncomfortable moments in the CSA.

- While examiners acknowledged the challenge of the exam for all candidates, there was widespread awareness of and sympathy for the particular challenges that IMG candidates face in the exam.

9.2.6   Educational interventions

- There is an important but 'hidden curriculum' which needs to be made explicit in educational interventions. Sociolinguistic analysis shows that much of this relates to fleeting and apparently mundane features which, cumulatively, can have large consequences for candidates.

- Any materials need to be based on real interactions (i.e. on the simulated cases in the CSA which are a real part of the licensing exam) and their use needs a new analytic

language which examiners, trainers and trainees can share.

- While some of the strategies of successful candidates can be taught, not all are susceptible to explicit teaching. It is important, therefore, that a habit of slowness and smallness is developed so that all groups can become more aware and self-critical.

9.3   Implications

Insights from this sociolinguistic research should contribute to better training and preparation and to a review of some aspects of the exam.

9.3.1   Training and preparation

- General practice is a talking/listening specialism. New training materials, based on sociolinguistic evidence are needed to reflect this. Understanding and developing good practice is notoriously difficult and paradoxical in this area, as this research has shown. These materials should be janus-faced – looking both ways, both inwards to those who design and run the exam as well as out to registrars and GP trainers.

- A new analytic language, illustrated in this report, is needed for talking about talk and interaction which is based on expertise, avoids emotional terms and provides information about some of the challenges facing IMGs of consulting in a new variety of English.

Candidates need materials and support to:

- Improve extended explanations, avoid and repair misunderstandings and be more aware of potential misalignments.

Materials also need to focus on:

- Aligning with patients and balancing this with institutional requirements, making talk more conversational, customising CSA phrases to sound less formulaic, convey their stance to the examiner through more explicit metacommunicating.

- There is a strong case for structuring aspects of training specifically around ethnically mixed groups of Associates in Training (AiTs) and arranging for 'buddy' practices so that AiTs can experience consulting with different patient populations. This would benefit all trainees.

9.3.2   Exam design and assessment implications

Those aspects of the exam that put particular weight on IMGs (the IPS domain in particular) should be reviewed. While some of the weight of the exam is an inevitable effect of moving to a different country for postgraduate practice with a different health service and consulting styles, our research can identify some ways in which some of these factors could be mitigated. This would involve some changes to the organisation of the CSA marking schedule. Seven years after setting up the CSA is an appropriate time to undergo a review and take the opportunity to

Conclusions and implications

look beyond current assumptions and evidential standards.

There should be a review of the role of communication and interpersonal skills in the CSA. Communication is assessed in data gathering and clinical management domains and it is essential that particular attention in marking is paid to how candidates (i) achieve clarity in explanations, management plans and other phases of candidate extended talk and ii) display accuracy and inferencing in listening to patients' talk. This review should also re-consider the role of IPS in the CSA for 4 reasons:

(i) the IPS is assessed both explicitly and implicitly and so carries more weight than the other two domains. This intense focus on a particular model of 'soft skills' of interpersonal relating is somewhat out of kilter with the changes in real world consulting as a result of the increasingly diverse patient population and use of computerised records.

(ii) there are aspects of consulting skills which the CSA cannot do well; most evidently, the assessment of interpersonal effectiveness in simulated encounters. The mix of subjectivity, simulation and standardisation produces a tension vulnerable to criticisms of unfairness. In particular, culturally specific, internal states such as empathy and rapport cannot be assessed by observers.

(iii) The extra implicit weighting falls most heavily on IMG candidates since this is the most linguistically and culturally demanding aspect of the CSA.

(iv) IPS/communication with linguistically diverse patient populations is not assessed. This means that IMGs' strengths in this area cannot be judged and UK graduates do not have to face the additional burden of interacting with RPs with a very different use of English from their own.

So, consideration should be given to redesigning the CSA marking schedules, to focussing on communication and alignment i.e. showing understanding, clear lines of communication and respect; and ensuring that that these are not over-assessed in comparison with data gathering and clinical management. An alternative might be to look elsewhere to consider other domains in a redesign in the marking schedule (for example RACGP 2008) This is a matter for the RCGP to decide.

Communication should also be judged, in a systematic way, as part of work-based assessment where candidates' skills in context are displayed. This could lead to the re-introduction of video recording real consultations where both the use of the EPR and candidates' skills in consulting in an increasingly linguistically diverse society could be tested. Simulated cases cannot do this effectively.

The RCGP is actively working to recruit more BME UK graduate and IMG examiners. The extent to which increasing these numbers significantly would have an impact on IMG failure rates is difficult to predict. (It would depend on how far any new examiners from this group drew on consulting styles which were somewhat different from the prevailing norms in the CSA).

However, there could be indirect benefits to the overall institutional environment of the exam and how it is perceived by IMGs. Continued efforts to recruit a more heterogeneous group of examiners remains a priority.

9.3.3 Limitations of the study

The mixed method approach combined a largely qualitative study with some quantitative linguistic analysis. The goal was to look at the detailed ways in which interactions are accomplished ie to put a microscope on the data in order to understand the differential pass rates in the exam. But this detailed analysis was also supported by looking across the data at patterns. A mixed method approach can make the methods vulnerable to criticism from both positivistic and interpretive traditions. Positivistic research expects large numbers and statistical conclusions and this research could only provide limited evidence within this paradigm. Interpretive micro-analysis looks intensively at small pieces of data for as complete an analysis as possible and again our data base of 40 cases was too extensive to allow for a complete analysis and so the focus was on aspects highlighted by previous studies.

The analysis of the 40 cases excluded any clinical omissions and errors by candidates since these fell outside the scope of the research. This raises questions about the relationship between poor performance overall and those who demonstrated clinical weaknesses. Finally, the number of examiner video feedback sessions was severely limited by the time constraints of the research and other limitations of ethnographic feedback studies. While further sociolinguistic research could address these limitations, this study has been able to look inside the CSA in new ways which are intended to stimulate creative thinking in addressing the challenges of institutional face to face assessments in an increasingly diverse society.

References

Aijmer, K. 2002. *English Discourse Particles. Evidence from a Corpus.* Amsterdam: John Benjamins.

Allan, K. 2013. 'Skilling the self: the communicability of immigrants as flexible labour'. In Duchêne et al (eds.): 56–78.

Alves de Lima, A., Conde, D., Costabel, J., Corso, J., & van der Vleuten, C. 2011. 'A laboratory study on the reliability estimations of the mini-CEX'. *Advances in Health Sciences Education.* On-line ahead of print. doi:10.1007/s10459-011-9343-y.

Balint, M. 1970. 'Quality, core values and the general practice consultation: issues of definition, measurement and delivery'. *Family Practice* 21(4): 458–468.

Bazzanella, C. 1990. 'Phatic connectives as interactional cues in contemporary spoken Italian'. *Journal of Pragmatics* 14(4): 629–647.

Berg, K., Blatt, B et al. 2014. 'Standardised patient assessment of medical student empathy: Ethnicity and gender effects in a multi-institutional study'. To appear in *Academic Medicine*.

Blackhall, L., Murphy, S., Frank, G., Michel, V. and Azen, S. 1995. 'Ethnicity and attitudes toward patient autonomy'. *Journal of American Medical Association* 274: 820–5.

Blatt, B., Spinazzi, N. and Greenburg, L. 2014. 'Communication skills training for resident physicians'. In Heidi Hamilton and Wen-ying Sylvia Chou (eds.) *The Routledge Handbook of Language and Health Communication.* London and New York: Routledge: 294–326.

Bleakley, A. 2003. 'Good and "poor" communication in an OSCE: education or training?' *Medical Education* 37 (3): 186–7.

Bolton, S. and Boyd, C. 2003. 'Trolley Dolly or Skilled Emotion Manager? Moving on from Hochschild's Managed Heart'. *Work Employment & Society* 17(2) 289–308.

Bonvicini, K., Perlin, M., Bylund, C., Carroll, G., Rouse, R. and Goldstein, M. 2009. 'Impact of communication training on physician expression of empathy in patient encounters'. *Patient Education and Counseling* 75: 3–10.

Boulet, J. R., McKinley, D. W., Norcini, J. J., & Whelan, G. P. 2002. 'Assessing the comparability of standardised patient and physician evaluations of clinical skills'. *Advances in Health Sciences Education: Theory and Practice* 7(2): 85–97.

Bouma, H. 2008. 'Is Empathy Necessary for the Practice of "Good" Medicine?' *The Open Ethics Journal* 2: 1–12 http://www.benthamscience.com/open/toj/articles/V002/1TOJ.pdf

Bourdieu, P. 1977. *Outline of a Theory of Practice* Cambridge: Cambridge University Press.

Bourdieu, P. 1990. 'Structures, habitus, practices'. In P. Bourdieu *The Logic of Practice*: 52–79. Stanford, CA: Stanford University Press.

Bourdieu, P. and Wacquant, L. 1992. *An Invitation to Reflexive Sociology.* Chicago: University of Chicago Press.

Britten N., Stevenson F., Barry C., Barber N., Bradley C. 2000. 'Misunderstanding in prescribing decisions in general practice: a qualitative study'. *British Medical Journal* 320:484–488.

Brown, B., Crawford, P. and Carter, R. 2006 *Evidence Based Health Communication.* Milton Keynes: Open University Press.

Brunsson, N. and Jacobsson. B. 2000. *A World of Standards.* Oxford: Oxford University Press.

Cameron, D. 2000. *Good To Talk: Living and Working in a Communication Culture*. London: Sage.

Campion, P., Foulkes, J., Neighbour, R. and Tate, P. 2002. 'Patient-centredness in the MRCGP video examination: analysis of a large cohort'. *British Medical Journal* 52: 691–2.

Carol, L. 1871 *Alice Through the Looking Glass.* London MacMillan and Co.

Carter, R and McCarthy, M. 1995. 'Spoken Grammar: What is it and how do we teach it?'. *ELT Journal* 49(3): 207–218.

Carter, R. 2004. *Language and creativity: the art of common talk.* London: Routledge.

Charon, R. 2001 'Narrative medicine: A model of empathy, reflection, professionalism and trust'. *JAMA* 286, 15: 1897–1902.

Cheng, D., Lew, R., Hershman, W. and Orlander, D. 2007. 'A cross-sectional measurement of medical student

empathy'. *Journal of General Internal Medicine* 22(10): 1434–8.

Clark, H. H. 1996. *Using Language*. Cambridge: Cambridge University Press.

Cook-Gumperz, J. 2001. 'Cooperation, collaboration and pleasure in work'. In A. Di Luzio, S. Günthner, S and F. Orletti (eds.) *Culture in Communication: Analyses of Intercultural Situations* Amsterdam: John Benjamins: 117–139.

Coupland, N., Giles, H. and Wiemann, J. (eds.) 1991. *'Miscommunication' and problematic talk.* Newbury Park, CA: Sage 374.

Crystal, D. 2007. *How Language Works.* London: Penguin.

De la Croix, A. and Skelton, J. 2009. 'The reality of role-play: interruptions and amount of talk in simulated consultations'. *Medical Education* 43(7): 695–703.

Denney, M. L., Freeman, A. and Wakeford, R. 2013 'MRCGP CSA: are the examiners biased, favouring their own by sex, ethnicity, and degree source?' *British Journal of General Practice* vol. 63 no. 616 e718-e725.

Derksen, F., Bensing, J. and Lagro-Jenssen, A. 2013. 'Effectiveness of empathy in general practice: a systematic review'. *British Journal of General Practice* 63/606: 76–84.

Drew, P. 1991. 'Asymmetries of knowledge in conversational interactions'. In I.Markova and K. Foppa (eds.) *Asymmetries in Dialogue.* London: Harvester Wheatsheaf: 21–48.

Duchêne, A., Moyer, M. and Roberts, C. (eds.) 2013. *Language, Migration and Social Inequalities.* Bristol: Multilingual Matters.

Epley, N. 2014. *Mind-wise: How we understand what others think, believe, feel and want*. London: Allen Lane.

Erickson, F. 1986. 'Listening and speaking'. In D. Tannen & J. Alatis (eds.), *Georgetown University Round Table on Languages and Linguistics*. Washington, DC: Georgetown University Press: 294–319.

Erickson, F. and Schultz, G. 1982. *The Counselor as Gatekeeper.* New York: Academic Press.

Erman, B. and Warren, B. 2000. 'The idiom principle and the open choice principle'. *Text and Talk* 20(1): 29–62.

Fairclough, N. 1992. *Discourse and Social Change.* Cambridge: Polity Press.

Ferguson N. 1977. 'Simultaneous speech, interruptions and dominance'. *British Journal of Social and Clinical Psychology* 16:295–302.

Forster, E.M. 1936. *A Passage to India.* London: Penguin Books.

Fox, R. and Lief, H. 1963. 'Training for "detached concern"'. In H. Lief (ed.) *The Psychological Basis of Medical Practice:* 12–35. New York: Harper and Row.

Frankel, R. 2009. 'Empathy research: a complex challenge'. *Patient Education and Counselling* 75: 1–2.

Fuertes, J. N., Gottdiener, W. H., Martin, H., Gilbert, T. C., & Giles, H. 2012. 'A meta-analysis of the effects of speakers' accents on interpersonal evaluations'. *European Journal of Social Psychology* 42: 120–133.

Giddens, A. 1991. *Modernity and Self-identity: self and society in the late modern age*. Cambridge: Polity Press.

Gluszek, A., & Dovidio, J. 2010. 'The way they speak: A social psychological perspective on the stigma of non-native accents in communication'. *Personality and Social Psychology Review,* 14: 214–237.

Goffman, E. 1959. *The Presentation of Self in Everyday Life*. London: Penguin.

Govaerts, M., van der Vleuten, C., Schuwirth, L., & Muijtjens, A. 2007. 'Broadening perspectives on clinical performance assessment: Rethinking the nature of in-training assessment'. *Advances in Health Sciences Education: Theory and Practice* 12(2), 239–260.

Greenhalgh, T. and Hurwitz, B. 1999. 'Why study narrative?' *British Medical Journal* 1999;318:48.

Grugulis, I. and Vincent, S. 2009. 'Whose skill is it anyway? "Soft" skills and polarisation'. *Work Employment & Society* 23 (4) 597–615.

Guardian 2014. http://www.theguardian.com/commentisfree/2014/mar/08/elderly-patronise-ageing-suits-nhs-respect-empathy

Gumperz, J. 1982. *Discourse Strategies.* Cambridge: Cambridge University Press.

Halpern, J. 2003. 'What is clinical empathy?' *J Gen Intern Med* 18(8): 670–674.

References

Hansen, K., Rakić, T. and Steffens, M. 2014. 'When actions speak louder than words: Preventing discrimination of nonstandard speakers'. *Journal of Language and Social Psychology* 33: 68.

Hellermann, J. and Vergun, A. 2007. 'Language which is not taught: The discourse marker use of beginning adult learners of English'. *Journal of Pragmatics* 39(1): 157–179.

Heritage, J. 1997. 'Conversation Analysis and Institutional Talk: Analysing data'. In D.Silverman (ed.) *Qualitative Analysis: Issues of Theory and Method.* London: Sage: 222–245.

Hinnenkamp, V. 2003. 'Misunderstandings: Interactional structure and strategic resource. In J. House, G. Kasper and Ross, S (eds.) *Misunderstandings in Social Life: Discourse approaches and problematic talk.* Harlow: Pearson: 57–81.

Hojat, M., Mangione, S., Nasca, T.J., Cohen, M., Gonnella, J.S., Erdmann, J.B., Veloski, J. & Magee, M. 2001. 'The Jefferson Scale of Physician Empathy: Development and Preliminary Psychometric Data'. *Educational and Psychological Measurement* April 2001 vol. 61 no. 2 349–365.

Hojat M., Gonnella J., Nasca T., et al. 2002 'Physician empathy: definition, components, measurement, and relationship to gender and specialty'. *American Journal of Psychiatry* 2002; 159 (9): 1563–1569.

Hojat, M., Mangione, S., Nasca, T., Rattner, S., Erdmann, J., Gonnella, J., & Magee, M. 2004. 'An empirical study of decline in empathy in medical school'. *Medical Education*, 38, 934–41.

Howie J., Heaney D. and Maxwell M. 2004. 'Quality, core values and the general practice consultation: issues of definition, measurement and delivery'. *Family Practice* 21(4): 458–468.

Jain, P. and Krieger, J. 2011. 'Moving beyond the language barrier: The communication strategies used by international medical graduates in intercultural medical encounters'. *Patient Education and Counseling* 84 (1): 98–104.

Kahneman, D. 2011. *Thinking Fast and Slow.* New York: Farrar, Strauss and Giroux.

Kanchandani, R. 2011. *The CSA: Views of trainers and GP registrars in Luton.* DVD.

Karpf, A. 2006. *The Human Voice: the story of a remarkable talent.* London: Bloomsbury.

Kaur-Bola, K. and Randhawa, G. 2012. 'Role of Islamic religious and cultural beliefs regarding intellectual impairment and service use: A South Asian parental perspective'. *Communication & Medicine* 9 (3): 241–251.

Kirch, M. 1979. Non-verbal communication across cultures'. *The Modern Language Journal* 416–423.

Kirklin, D. 2001. 'Metaphors for medicine: revealing reflections or just popular parodies?' *Journal of Medical Humanities* 27: 89.

Koenig A. and Eagly, A. 2005. Stereotype threat in men on a test of social sensitivity. *Sex Roles* 52 (7/8): 489–495.

Kogan, J., Holmboe, E. and Hauer, K. 2009. 'Tools for direct observation and assessment of clinical skills in medical trainees: A systematic review'. *Journal of the American Medical Association* 302:13–16.

Konijin, E. 1997. *Acting Emotions: Shaping Emotions on Stage.* Amsterdam: Amsterdam University Press.

Kring, A. and Gordon, A. 1998. 'Sex differences in emotion: Expression, experience, and physiology'. *Journal of Personality and Social Psychology*, Vol 74(3): 686–703.

Kurtz, S., Silverman, J. and Draper, J. 1998 . *Teaching and learning communication skills in medicine.* Abingdon: Radcliffe Medical Press.

Kurtz, S. and Silverman, J. 1996. *The Calgary-Cambridge Observation Guides. Medical Education 30: 83 – 8.9.*

Labov, W. and Waletzky, J. 1967. 'Narrative analysis'. In J. Helm (ed.) *Essays on the Verbal and Visual Arts.* Seattle: U. of Washington Press: 12–44.

Lakoff, G. 1987. *Women, fire, and dangerous things: What categories tell us about the mind.* Chicago: University of Chicago.

Lambert, W., Hodgson, R., Gardner, R. and Fillenbaum, S. 1960. 'Evaluational reactions to spoken language'. *Journal of Abnormal and Social Psychology,* 60, 44–51.

Larson, E., Yao, X. 2005. Clinical empathy as emotional labour in the patient-physician relationship. *JAMA,* 293, 1100–1106.

Lindemann, S. and Subtirelu, N. 2013. 'Reliably Biased: The Role of Listener Expectation in the Perception of Second

Language Speech'. *Language Learning 63(3)* 567–594.

Lingard, L., Albert, M and Levinson, W. 2008. 'Grounded theory, mixed methods, and action research'. *BMJ* 2008;337:a567 doi:10.1136/bmj.39602.690162.47.

Little, P., and Everitt, H., Williamson, I., Warner, G., Moore, M., Gould, C. et al. 2001. 'Preferences of patients for patient centred approach to consultation in primary care: observational study'. *British Medical Journal* 322:468–472.

McCarthy, M. and Carter, R. 1994. *Language as Discourse: Perspectives for Language Teaching.* London: Longman.

McKinstry B. 2000. 'Do patients wish to be involved in decision making in the consultation? A cross sectional survey with video vignettes'. *British Medical Journal* 321:867–71.

McManus, I. and Wakeford, R. 2014. 'PLAB and UK graduates' performance on MRCP(UK) and MRCGP examinations: data linkage study'. *BMJ* 2014;348:g2621.

McWhinney, B. 2013. CLAN programme http://childes.psy.cmu.edu/manuals/CLAN.pdf

Marshall, R. and Bleakley A. 2009. 'The death of Hector: pity in Homer, empathy in medical education'. *Med Humanities* 2009;35:7e12.

Mercer, S. and Reynolds, W. 2002. 'Empathy and quality of care'. *BJGP* vol. 52 no. Suppl S9-12.

Mercer, S., Maxwell, M., Heaney, D. and Watt, C. 2004. 'The consultation and relational empathy (CARE) measure: development and preliminary validation and reliability of an empathy based consultation process measure. *Family Practice* : 21(6): 699–705.

Milroy, L. and Milroy, J. 1992. 'Social network and social class: Toward an integrated sociolinguistic model'. *Journal of Language in Society* 21(1) 1–26.

Miskovic-Lukovic, M. 2009. 'Is there a chance that I might kinda sort of take you out to dinner?: The role of the pragmatic particles 'kind of' and 'sort of' in utterance interpretation'. *Journal of Pragmatics* 41(3):602–625.

Mohanna, K. 2011. 'Exploring the Royal College of General Practitioners' Clinical Skills Assessment'. Unpublished thesis is partial completion of EdD. Institute of Education, University of London.

Moss, B. and Roberts, C. 2005. 'Explanations, Explanations, Explanations: how do patients with limited English construct narrative accounts in multi-lingual, multi-ethnic settings and how can GPs interpret them?' *Family Practice* 4(3): 1–7.

Neighbour, R. 1987. *The Inner Consultation – how to develop an effective and intuitive consulting style.* Lancaster: Kluwer Academic Press.

O'Grady, C. 2011. 'The nature of expert communication as required for the General Practice of medicine – A discourse analytical study'. Unpublished PhD thesis, Department of Linguistics, Macquarie University.

Pandya, H. 2011. Professionalism and Communication in medical practice. *Medicine Update 2011.*

Pearce, C., Dwan, K., Arnold, M., Phillips, C., & Trumble, S. 2009. 'Doctor, patient and computer – a framework for the new consultation'. *International Journal of Medical Informatics* 78: 32–38.

Pearce, C., Trumble, S., Arnold, M., Dwan, K., & Phillips, C. 2008. 'Computers in the new consultation: within the first minute'. *Family Practice, 2*5(3):202–208.

Pedersen R. 2009. 'Empirical research on empathy in medicine: critical review'. *Patient Education and Counseling* 76:307e22.

Pickering, M., and Garrod, S. 2004. 'Toward a mechanistic psychology of dialogue'. *Behavioral and Brain Sciences* 27 (2): 169–225.

Porter, R. 2003. *Flesh in the Age of Reason.* London: Allen Lane.

Rakić, T., Steffens, M. C., and Mummendey, A. 2011. 'Blinded by the accent! The minor role of looks in ethnic categorization'. *Journal of Personality and Social Psychology* 100: 16–29.

Reay, D., David, M. and Ball. S. 2005. *Degrees of Choice: Class, Gender and Race in HE choice process.* London: Trentham Books.

Roberts, C and Campbell, S 2006 'Talk on Trial: Job interviews, language and ethnicity' (2006) *Research Report No. 344* DWP http://www.dwp.gov.uk/asd/asd5/rrs2006.asp#talkonntrial

Roberts, C., Sarangi, S., L. Southgate, L., R. Wakeford, R., and Wass, V. 2000. 'Oral examinations – equal

References

opportunities, ethnicity and fairness in the MRCGP'. *British Medical Journal 320*: 370–374.

Roberts, C., Sarangi, S. and Moss, B. 2004. 'Presentation of self and symptoms in primary care consultations involving patients from non-English speaking backgrounds'. (2004) *Communication and Medicine Vol 1/2*, pp.159–170.

Roberts, C., Wass, V., Jones, R., Sarangi, S., and Gillett, A. 2003. 'A discourse analysis study of "good" and "poor" communication in an OSCE: A proposed new framework for teaching students'. *Medical Education 37*, 192–201.

Roberts, C., Moss, B., Wass, V., Sarangi, S. and Jones, R. 2005. 'Misunderstandings: a qualitative study of primary care consultations in multilingual settings, and educational implications'. *Medical Education 39*: 465–475.

Romero Trillo, J. 2002. 'The pragmatic fossilisation of discourse markers in non-native speakers of English'. *Journal of Pragmatics* 34(6):769–784.

Roter D. 1989. 'Which facets of communication have strong effects on outcome – A meta-analysis'. In M. Stewart and D. Roter, (eds). *Communicating With Medical Patients*. Newbury Park: Sage Publications: 183–196.

Royal Australian College of General Practitioners. [2005] 2008. *The college examination. A handbook for candidates and examiners.* Melbourne: RACGP.

Royal College of General Practitioners. 1972. 'The future general practitioner: learning and teaching'. London: RCGP.

Royal College of General Practitioners. 2013. Annual Report 2012-13. London: RCGP.

Ruusuvuori, J. 2007. 'Managing affect: integration of "empathy" and problem solving in health care encounters'. *Discourse Studies* 9(5): 597–622.

Sacks, H. 1992. *Lectures on Conversation.* Volume I and II. G.Jefferson (ed.). Cambridge: Blackwell.

Sanden I., Linell, P., Starkhammar, H. and Larsson U.S. 2001. 'Routinisation and Sensitivity: Interaction in Oncological Follow-Up Consultations'. *Health* 5(2):139–163.

Schegloff, E. 1987. 'Some sources of understanding in talk-in-interaction'. *Linguistics* 25(1): 201–18.

Schiffrin, D. 1987. *Discourse Markers.* Cambridge: Cambridge University Press.

Schmitt, N. and Carter, R. 2004. 'Formulaic Sequences in Action: An Introduction'. In N. Schmitt, (ed.) *Formulaic Sequences: Acquisition, Processing and Use*. Amsterdam: John Benjamins. 1–22.

Schouten, B., Meeuwesen, L. 2006. 'Cultural differences in medical communication: a review of the literature'. *Patient Education and Counseling* 64: 21–34.

Schouten, B., Meeuwesen, L. and Harmesan, H. 2009. 'GPs' Interactional styles in consultations with Dutch and Ethnic minority patients'. *Journal of Immigrant Minority Health* 11: 468–475.

Seale, C. (ed.) 2004. *Researching Society and Culture.* London: Sage.

Seale, C., Butler, C., Hutchby, I., Kinnersley, P. and Rollnick, S. 2007. 'Negotiating frame ambiguity: A study of simulated encounters in medical education. *Communication and Medicine* 4(2): 177–87.

Sennett, R. 2008. *The Craftsman.* London: Allen Lane.

Smajdor, A., Stöckl, A. and Salter C. 2011. 'The limits of empathy: problems in medical education and practice'. *Journal of Medical Ethics.* 2011;37:380–383.

Stenström A-B. 1994. *An introduction to spoken interaction.* London: Longman.

Stepien, K.A. and Baernstein, A. 2006. 'Educating for empathy: a review'. *Journal of General Internal Medicine* 21(5): 524–530.

Stewart, M., 2001. 'Towards a global definition of patient-centred care'. *British Medical Journal* 322: 444–5.

Stockdale, M., and Warelow, P. 2000. 'Is the complexity of care a paradox?' *Journal of Advanced Nursing,* 31, 1258–1264.

Stokes, R. and Hewitt, J. 1977. 'Aligning Actions'. *American Sociological Review* 41/5: 838–849.

Swinglehurst, D., Greenhalgh, T., and Roberts, C. 2012. 'Computer templates in chronic disease management: ethnographic case study in general practice'. *BMJ Open: 2:e001754.doi:10.1136/bmjopen-2012-001754.*

Swinglehurst, D., Roberts, C, Li S et al. 2014 'Beyond the "dyad": qualitative re-evaluation of the changing clinical

consultation'. *BMJ Open* 2014;4:e006017.

Tomorrow's Doctors 2009. London: General Medical Council.

UK Census for England and Wales. 2011. Office of National Statistics.

Walsh, I.P. 2008. '"Whose voice is it anyway?" Hushing and hearing "voices" in speech and language therapy interactions with people with chronic schizophrenia'. *International Journal of Language and Communication Disorders* 43, (supp 1): 81–95.

Walsh, I.P. and Kovarsky, D. 2011. 'Establishing relationships in speech and language therapy when working alongside people with mental health disorders'. In R. Fourie (ed.) *Therapeutic Processes for Communication Disorders*. London: Psychology Press: 123–137.

Wass, V., Roberts, C., Hoogenbaum, R., Jones, R. and Van der Vleuten, C. 2003. 'Effect of ethnicity on performance in a final OSCE examination: qualitative and quantitative study'. *British Medical Journal* 326: 800–803.

Woolf, K., Potts, H. and McManus, I. 2011. 'Ethnicity and academic performance in UK trained doctors and medical students: systematic review and meta-analysis'. *BMJ* 2011;342:d901.

Wray, A. 2002. *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

West, C. 1998. 'When the Doctor is a "Lady": Power, status and gender in physician-patient encounters'. In J. Coates (ed.) *Language and Gender: A Reader*. Oxford: Blackwell Publishing: 396–412.

Wittgenstein, L. 2001 [1953]. *Philosophical Investigations.* Oxford: Blackwell Publishing.

Yeates, P., O'Neill, P., Mann, K. and Eva, K. 2013. 'Seeing the same thing differently: Mechanisms that contribute to assessor differences in directly-observed performance assessments'. *Advances in Health Science Education* 2013 18:325–341.