

*Final Report For The Occupational English Test Centre*

# **Towards improved healthcare communication**

*Development and validation of language  
proficiency standards for non-native English  
speaking health professionals*

**August 2013**

*Catherine Elder  
Tim McNamara  
Robyn Woodward-Kron  
Elizabeth Manias  
Geoff McColl  
Gillian Webb  
John Pill*



The *Towards Improved Healthcare Communication* project was carried out with funding from the Australian Research Council Linkage Project LP0991153 and the Occupational English Test (OET) Centre (Industry Partner).



**APAI Scholar and Project Manager**

John Pill

**Associate Researchers**

Sally O'Hagan

Diana van Die

**Acknowledgments:** The research team gratefully acknowledges the contributions of the staff of The OET Centre and of the OET assessors who participated in Phase Two of the project. We would also like to thank the medical, nursing and physiotherapy health professionals and health professional educators who participated in the workshops in Phases One and Three.

We are indebted to the patients who allowed us to audio-record their interactions with health professionals and to the students and supervisors who consented to us audio-recording their feedback sessions. Our thanks are also owed to the Project Advisory Team, who provided valuable feedback and recommendations throughout the project.

Finally, we would like to acknowledge the intellectual input of our colleagues in the Language Testing Research Centre from conceptualisation to completion of the project.

## TABLE OF CONTENTS

<b>Executive summary and recommendations</b>	<b>5</b>
<b>0. Introduction</b>	<b>9</b>
<b>1. Background to the project</b>	<b>9</b>
<b>2. OET: A health-specific test of workplace communication</b>	<b>9</b>
<b>3. Study aims and research questions</b>	<b>11</b>
<b>4. Approach and training</b>	<b>12</b>
<b>5. Phase One: Establishing professionally relevant criteria</b>	<b>13</b>
<i>5.1 Findings</i>	<i>15</i>
<i>5.2 A model of what is valued by doctors in the doctor-patient consultation</i>	<i>18</i>
<i>5.3 Summary of Phase One</i>	<i>19</i>
<b>6. Phase Two: Application of professionally relevant criteria to OET</b>	<b>19</b>
<i>6.1 Themes for inclusion in an expanded OET construct</i>	<i>20</i>
<i>6.2 Operationalising themes as criteria</i>	<i>21</i>
<i>6.3 Assessor training</i>	<i>23</i>
<i>6.4 Functioning of new criteria</i>	<i>24</i>
<i>6.5 Summary of Phase Two</i>	<i>27</i>
<b>7. Phase Three: Setting minimum standards for professional registration</b>	<b>27</b>
<i>7.1 Data elicitation</i>	<i>28</i>
<i>7.2 Findings</i>	<i>29</i>
<i>7.3 Summary of Phase Three</i>	<i>30</i>
<b>8. Recommendations for practice</b>	<b>30</b>
<i>8.1 Changes to OET</i>	<i>30</i>
<i>8.2 Training requirements</i>	<i>31</i>
<i>8.3 Consultation and dissemination</i>	<i>32</i>
<b>9. Recommendations for further research</b>	<b>32</b>
<i>9.1 Validating the new criteria with additional professions</i>	<i>33</i>
<i>9.2 Setting cut-scores with additional professions</i>	<i>33</i>
<i>9.3 Exploring the role of non-verbal behaviour</i>	<i>33</i>
<i>9.4 Understanding a wider range of interactions in healthcare</i>	<i>33</i>
<i>9.5 Written communication in healthcare settings</i>	<i>34</i>

<b>References</b>	<b>35</b>
<b>Appendix 1: Project team</b>	<b>39</b>
<b>Appendix 2: Reference group members</b>	<b>40</b>

## **EXECUTIVE SUMMARY AND RECOMMENDATIONS**

This report describes the aims, methods and outcomes of a cross-disciplinary project designed to establish more relevant and rigorous English communication standards for non-native English speaking health professionals applying to practise their profession in Australia. The project was funded under the Australian Research Council Linkage scheme with matching funds from the Occupational English Test Centre, the partner organisation. It was triggered by concerns that current English language screening procedures for overseas-trained health professionals might be insufficient for the task of attesting adequate language skills for effective functioning in the workplace.

The particular focus of the project was on the criteria used for assessment on Occupational English Test (OET), accepted by twelve professions as a language screening tool to ensure that overseas-trained applicants seeking to practise their profession in Australia meet legislated minimum English proficiency requirements. The particular focus of the project was on speaking, a critical skill for the workplace, and on whether the criteria currently used to assess performance on the speaking sub-test were sufficiently aligned with what health professionals perceive to be important for effective communication in the workplace. The project was conducted in three phases over a three-and-a-half year period, drawing on the expertise of applied linguistics on the one hand and health professional experts from three of the twelve health professions recognising OET (Medicine, Nursing and Physiotherapy) on the other.

The study had three aims:

- (1) to probe health professionals' views of spoken communication by eliciting judgements from health professionals in work-related teaching and assessment contexts (for Medicine, Nursing and Physiotherapy) and thereby identify valued aspects of professional communication (Phase One).
- (2) to translate these valued aspects of communication into professionally relevant criteria that could be applied to the assessment of migrant health professionals on OET and to explore how well these criteria functioned in measurement terms (Phase Two).

- (3) to involve the three groups of health professionals in making decisions about minimum passing standards for entry into each health profession by inviting their judgements on the adequacy of a representative sample of recorded speaking performances on OET (Phase Three).

The methodology used to address these aims is described in detail in the report which follows. The outcomes of the study's three phases are summarised below.

**Phase One:** The research elicited feedback from health professionals from Medicine, Nursing and Physiotherapy on samples of professional interaction in simulated or workplace contexts. Analysis of their feedback yielded a conceptual model representing the aspects of a health professional's performance in the health professional–patient consultation that are valued by expert practitioners (Pill, forthcoming). The model comprises three mutually dependent skill sets: *Communication skills*, *Clinical skills*, and *Practitioner skills* which draw on a repertoire of *Interactional tools* as required in performance.

**Phase Two:** The model allowed us to identify valued aspects of communication that were amenable to inclusion on a language test. These included not only some of the language skills covered by the current OET assessment criteria but additional aspects. These additional aspects, drawn from the *Communication skills* and *Interactional tools* components of the model, were translated into two professionally–relevant criteria to be added to the current linguistic criteria on OET: CLINICIAN ENGAGEMENT and MANAGEMENT OF INTERACTION. OET language assessors were trained in the application of these new criteria using a checklist of indicators to assist them in recognising relevant features of behaviour. These new criteria were found to be useable by the assessors and to function well in measurement terms.

**Phase Three:** The three groups of health professionals participating in standard–setting sessions using the new criteria were able to classify OET candidates' performances according to their perceived readiness to cope with spoken communicative tasks in the clinical setting and they showed acceptable levels of agreement in their judgements. Their collective classifications were used to establish new cut–scores on OET. The estimated impact of these new cut–scores appeared to vary somewhat across

professions, resulting in slightly lower pass rates for two of the three professions, but at the same time a greater number of candidates being classified with A (highest) rather than B (minimum passing) grades, confirming that they were well-equipped for successful communication in clinical settings.

The study's findings have implications both for practice in relation to the assessment of overseas-trained health professionals and for research. These implications are framed in the form of recommendations below.

Practical recommendations can be divided into three major areas. These include changes to OET, training requirements, consultation and dissemination of outcomes.

Regarding *changes to OET*, it is recommended that:

- the existing linguistic criteria of the OET speaking sub-test be maintained given their perceived importance to the health profession;
- the current OVERALL COMMUNICATIVE EFFECTIVENESS criterion of the OET speaking sub-test be replaced by the two more explicit and professionally relevant criteria derived from the current study, CLINICIAN ENGAGEMENT and MANAGEMENT OF INTERACTION;
- consideration be given to how speaking performance on OET is reported—whether as a single overall score or as a profile with both a linguistic and a more professionally-/clinically-relevant component; and
- cut-scores (and by implication passing standards) on OET be examined for each profession in light of results of the standard-setting component of this study.

Regarding the *training of OET assessors*, it is recommended that:

- the checklist developed for the current study be used to guide OET language assessors in the application of the new criteria (if adopted); and
- workshops be conducted for the training of OET assessors following the model adopted in this study.

Regarding *consultation and dissemination*, it is recommended that:

- health professional boards and other representative bodies for the 12 professions catered for by OET be apprised of the outcomes of the current study;
- these bodies be asked to consider the implications of proposed changes to OET, how such changes could be implemented and evaluated, and how test scores should be reported; and
- the checklist devised for the training of assessors in the new criteria be made available to language teachers and test-takers to familiarise them with the new communicative demands of the test and (by implication) with the expectations of Australian healthcare work settings.

Recommendations for further research are proposed in the interests of consolidating and enhancing the validity of OET. Accordingly it is recommended that:

- the new professionally relevant criteria be trialled and validated with the nine professions served by OET but not included in the current study;
- additional standard-setting workshops following the procedures used for this study be conducted with these professions for the purpose of setting new cut-scores;
- the impact of including non-verbal behaviours amongst the aspects of communication assessed by OET be explored, using a sample of video-recorded rather than audio-recorded OET role-plays; and
- a study of the discourse demands of inter- and intra-professional communication be conducted with a view to possibly expanding the scope of interactions currently assessed on OET.

The findings of the current research on spoken communication are soon to be complemented by a study of written communication practices in healthcare settings, for which the research team has recently obtained ARC Linkage funding (LP130100171). Research in this and the other areas proposed above will offer new insights into the nature of healthcare communication and inform the ongoing renewal of OET, enhancing public understanding of the test and confidence in its predictive and diagnostic power.

## **0. INTRODUCTION**

This report describes a project funded in 2009 under the Australian Research Council Linkage scheme with the aim of establishing relevant and rigorous English communication standards for non–native English speaking health professionals applying to practise their profession in Australia. The project brought together 1) a team of University of Melbourne academics including language testing experts, medical, nursing and physiotherapy educators and applied linguists with expertise in healthcare communication (Appendix 1) and 2) an English language testing agency, the Occupational English Test Centre (now jointly owned by Box Hill Institute and Cambridge English Language Assessment). It is anticipated that the outcomes of the project reported here will result in more principled decision–making about professional readiness in the sphere of communication, thereby ensuring that health professionals are better equipped to communicate with their patients in the workplace.

## **1. BACKGROUND TO THE PROJECT**

Concern has been expressed in Australia about the English entry standards required for effective functioning in healthcare settings (Birrell, Rapson & Smith, 2005), given the impact of healthcare communication on the quality of healthcare delivery and outcomes (Birrell & Schwartz, 2006; Hawthorne & Birrell, 2002). In particular, the limited language proficiency of non–native health professionals, who play a crucial role in meeting Australia’s health workforce shortages (Australian Government Department of Health and Ageing, 2008; Barton, Hawthorne, Singh & Little, 2003; Douglas, 2008), has been identified as a potential obstacle to effective communication and decreased quality of care (Eggly, Musial & Smulowitz, 1999; House of Representatives Standing Committee on Health and Ageing, 2012; McDonnell & Usherwood, 2008; Xu, 2010). With substantial proportions of overseas–trained health professionals in the workforce (Hawthorne, 2012), these communication issues need to be systematically addressed.

## **2. OET: A HEALTH–SPECIFIC TEST OF WORKPLACE COMMUNICATION**

The Occupational English Test (OET), one of the best known tests of language for specific purposes (LSP) (Douglas, 2000), was designed to establish the adequacy of the workplace related communication skills of migrant health professionals whose training was not done in English (McNamara, 1996). Administered by The OET Centre, the

partner organisation for this project, OET is held in major cities in Australia and worldwide and is recognised by twelve different health professions, with doctors and nurses forming the largest test-taker groups and physiotherapists mandating a pass on this test for all non-native applicants for entry to the profession.

OET was originally developed under contract to the Australian Federal Government to meet the requirement within current legislation for the assessment of language skills separately from the assessment of other aspects of clinical competence. The issues raised by this legislative separation are at the heart of this project, and correspond to a central theoretical issue in the testing of second language proficiency within specific work-related contexts. Tests such as OET, developed primarily by linguists, claim to act as proxies for the demands of the communicative settings faced by candidates when they enter the workplace. OET was designed to replicate the critical tasks of the workplace setting (McNamara, 1996) and measure candidates' abilities across the skills of listening, reading, writing and speaking, in simulated workplace contexts. Speaking, generally agreed to be the most critical skill in healthcare communication, was the focus of the current study. The speaking tasks, specific to each of the professions taking the test, involve role-plays of typical health-related scenarios involving clinicians and patients or carers, developed collaboratively with healthcare experts, and are designed to mirror the communicative requirements of each health profession. Performance on these *test* tasks is then used as the basis for inferences about the performance of the candidate in the *non-test setting*. Test performances on OET have, however, been rated against a common set of essentially *linguistic* criteria developed by McNamara (1996) without input from health professionals—a contentious issue within the study of LSP testing (Basturkmen & Elder, 2004; Douglas, 2001; Douglas & Myers, 2000; Jacoby & McNamara, 1999). These criteria, although applied by highly trained language specialists, whose judgements are carefully monitored and statistically calibrated, may nevertheless differ from those which end-users of test scores (in this case, health professionals) see as critical to effective healthcare communication—a gap which has been noted in other LSP studies (e.g., Brown, 1995; Dias, Freedman, Medway & Paré, 1999; Elder, 1993, 2001; McNamara, 1996). Moreover, the appropriateness of the *passing standards* on OET required in terms of these criteria (usually a B grade for each skill area; this is set by the relevant health board) is the subject of ongoing debate, especially among those with experience of the clinical communication of health

professionals who have passed the test and may nevertheless struggle to cope in clinical settings, or candidates who feel that they have been unreasonably excluded from the workplace on the basis of performance on the test. Given that the pass standards were initially established without a formal standard setting procedure (e.g., following procedures described in Cizek & Bunch, 2007; Cizek, Bunch & Koons, 2004; Zieky, 2001), and that such tests bear a heavy burden of responsibility for making decisions affecting the careers of candidates and the people for whom they are subsequently responsible, the claim of the test to form a valid basis for determining professional entry demanded further investigation. The project described in this report was designed to validate both the *criteria* against which candidates are assessed, and the performance *standards* required for admission to clinical settings.

### **3. STUDY AIMS AND RESEARCH QUESTIONS**

The *first* aim of the study was to probe health professionals' views of spoken communication, drawing on the notion of "indigenous assessment" (Jacoby, 1998; Jacoby & McNamara, 1999). This was done by eliciting judgements from health professionals in work-related teaching and assessment contexts independent of the test setting (i.e. via training materials involving professional interaction or in training sites where professional interaction is routinely evaluated). Professionally relevant criteria were then formulated for the three professions (Medicine, Nursing and Physiotherapy) represented in the study.

The *second* aim of the study was to train language experts in the use of these professionally relevant criteria to assess the speaking performances of migrant health professionals on OET and to explore how well these criteria functioned in measurement terms.

The *third* aim of the study was to involve the three groups of health professionals in making decisions about minimum passing standards for entry to each health profession by inviting them to make judgements about the adequacy or otherwise of a representative sample of recorded speaking performances on OET.

The following research questions were addressed in the study:

- (1) What criteria underlie *health professionals'* judgements of the spoken clinical communication of non–native English speaking health professionals in routine assessment situations within the professional context?
- (2) Can such professionally relevant criteria be used as the basis for language assessments carried out by *language experts* of migrant health professionals seeking registration in Australia?
- (3) What *minimum standards* in terms of these criteria should be set for professional registration of migrant health professionals?

#### **4. APPROACH AND TRAINING**

The research questions outlined above were addressed over a three–year period and involved three key stakeholder groups: doctors, nurses and physiotherapists. The three–year timeframe offered an opportunity for the appointment of Mr John Pill, a funded scholar supported by an Australian Postgraduate Award (Industry) (APAI), with Master's level qualifications in applied linguistics/language testing, to complete a related doctoral project under the supervision of the multidisciplinary research team.

The project was carried out with the assistance of a Project Reference Group (see Appendix 2), involving representatives from three professions, to assist the research team and the APAI scholar with access to relevant professional sites and with the selection and recruitment of a representative sample of participants.

The project was conducted in three phases as summarised below in Table 1, with an extension granted for the completion of the PhD thesis and writing up of research and publications. More details of each phase and its outcomes are provided in the following sections of the report.

Table 1. Project phases and timeline

Phase	Timeline	Aim	Outcomes
<b>ONE</b> (Research Question 1)	Oct 2009 – Dec 2010	Establish professionally relevant criteria for assessing clinical communication skills of non-native speakers of English	<ul style="list-style-type: none"> <li>An empirically based model representing the aspects of a health professional's performance in the health professional-patient consultation that are valued by expert practitioners (Pill, forthcoming).</li> </ul>
<b>TWO</b> (Research Question 2)	Jan 2011 – Jan 2012	Apply these criteria in the test of English language skills required for professional registration of immigrant health professionals (OET)	<ul style="list-style-type: none"> <li>Revised criteria for the OET speaking test to incorporate two new criteria based on the outcomes of Phase One.</li> <li>A piloted methodology for training OET assessors to use the new criteria.</li> </ul>
<b>THREE</b> (Research Question 3)	Feb 2012 – Sep 2012	Set minimum standards on OET for professional registration of immigrant health professionals	<ul style="list-style-type: none"> <li>Minimum pass standards set by health professional using the revised criteria.</li> </ul>
<b>WRAP-UP</b>	Jan 2013 – Jun 2013	Report findings	<ul style="list-style-type: none"> <li>Doctoral dissertation (submitted August 2013).</li> <li>Final report (August 2013).</li> </ul>

## 5. PHASE ONE: ESTABLISHING PROFESSIONALLY RELEVANT CRITERIA

Phase One of the project investigated the views of three groups of expert health professionals, that is, doctors, nurses and physiotherapists, on what constitutes effective clinical communication. We sought to elicit the indigenous criteria of the expert health professionals, in this case, clinical educators, from a variety of settings, namely in workshops with a video stimulus of actual or simulated patient-trainee interactions, from written feedback on routine general practice trainee-patient interactions, and live interactions with trainees and patients in a clinic/ward setting. An overview of these settings and data is provided in Table 2.

Table 2. Overview of stimulus and data for the three professions in Phase One

STIMULUS	Videod interaction – trainee & patient/ simulated patient, trainee setting	Routine consultations— general practice (GP) trainees and a series of patients—GP setting	Live interaction—trainee and patient. Clinic/ward setting
DATA	Educators' commentary elicited at workshop (audio) n=33 educators, 7 workshops	Educators' feedback report to trainee (written) n=92 reports	Supervisors' feedback to trainee <i>in situ</i> (audio) n=11 supervisors n=16 feedback episodes

For each of the three professions being investigated, two to three workshops were held. At each profession-specific workshop, a group of up to eight educators was shown two video recordings of trainee-patient interactions. After viewing the interactions, the health professionals' (HPs') oral feedback was elicited; this feedback constituted the data for the research and was analysed thematically. The workshop setting allowed for discussion between participants. The video stimulus included non-native speaker international medical graduate (IMG) trainee doctor-patient interactions, native speaker physiotherapy trainee-patient interactions, non-native speaker physiotherapy trainee-patient interactions, and native speaker trainee nurse-patient interactions respectively. The general practice written reports provided feedback on the performance of both IMG trainees (n=46) as well as Australian medical graduate (AMG) trainees (n=46) interacting with patients as part of routine GP consultations. The physiotherapy live interactions were in a variety of patient care settings including outpatients and rehabilitation. Feedback on fourth-year physiotherapy student interactions with patients from clinical supervisors was captured *in situ*. The physiotherapy students were all native or near native speakers of English. For a detailed description of the study design for Phase One, including participant recruitment rationale and findings, see Elder et al. (2012) for the workshops, Pill (forthcoming) for the written reports and medicine workshops, Woodward-Kron et al. (2012) for the physiotherapy workshops and clinic/ward settings, and O'Hagan et al. (in press) for the nursing workshops.

A thematic content analysis was undertaken of all the transcribed data. Themes were identified based on components of trainee/student performance that emerged from the medical workshop data in the first instance. Over a series of meetings, the draft scheme was presented to other project members for discussion and amendment. It was then applied to the workshop data from all three professions by three coders working independently with each profession. Through a process of refining the existing categories, and adding further categories, the final coding scheme was developed and applied to all the workshop data, the written report data, and the supervisor *in situ* feedback data.

## 5.1 Findings

In response to the question of what criteria underlie health professionals' judgements of the spoken clinical communication of non-native English speaking health professionals in routine assessment situations within the professional context, twenty-one inter-related themes were identified from the workshop, written report, and live interaction trainee-patient data. The themes with definitions and examples are provided in Table 3. Both positive [+] examples as well as negative [-] realisations are included. A super-ordinate category (indicated in the Description column of Table 3) draws together aspects of other themes. Indentation (highlighted using "...") indicates membership of a broader category in the thematic hierarchy. The broader category is above the indented theme in the table. A theme label in **bold type** indicates that the term is used by participants in the data.

As Table 3 shows, there are a number of super-ordinate categories. The function of the super-ordinate categories *Patient-centredness* and *Efficiency* is to do with achieving the goals of the consultation. The super-ordinate categories of *Communication skills*, *Clinical skills* and *Practitioner skills* have to do with the performance of the consultation. The foundational elements are *Knowledge*, *World view*, *Affect* and *Language*. These latter four aspects are the fundamental resources and influences that the health professional brings to the consultation.

How the themes are interrelated and combine to achieve the work of the consultation is discussed in the following section.

Table 3. Themes identified as important in health professionals' feedback

Theme	Code	Description and Example
<b>Patient-centredness</b>	PC	attitude/approach to patient; awareness of/sensitivity to patient's ideas, concerns, expectations and situation you negotiated with the patients in a way that empowered them, eg, when you were talking about what to do about the first woman's tummy pain, you said "we can decide together". [Med Report] [+]
<b>Efficiency</b>	EF	capacity to manage available resources, including time and effort, to complete the intended task There was a lot to cover with this patient and you were efficient and methodical in your consultation structure [Med Report] [+]

Interactional tools	IT	interactional techniques/resources In the same way that you are currently signposting what examination you would like to perform, it can be helpful to signpost the fact that you are going to use the computer [Med Report] [+/-]
... Terminology	TE	use of particular words and phrases, technical jargon using the word "illicit" for drugs which we tend to sort of say "recreational" [Med Wkshp] [-]
<b>Communication skills</b>	CS	super-ordinate category Basically her communication skills were good [Med Wkshp] [+]
... Manner	MA	behaviours towards patient of engagement, rapport, self-presentation; verbal encouragement Everyone is busy. It's to be un-busy, and to look at that patient and go, "What is it that that patient really needs?" and make them feel like at that moment they are the most important person when you're talking to them. [Nursing Wkshp] [-]
... Non-verbal communication	NV	eye contact, facial expression, posture and movement of head, limbs and body; non-verbal sounds his eye contact and body language was reasonably good [Physio Wkshp] [+]
<b>Clinical skills</b>	CL	super-ordinate category Excellent clinical skills. Your history taking is good but could be a bit more thorough and your examinations are good. [Med Report] [+]
... Content	CO	knowledge-in-interaction; subject knowledge, clinical and procedural knowledge as realised in the consultation; coverage of the issues she said "okay, you've had this stroke," didn't actually ask about his current function and what effects it's had and um change pre- to post morbid [Physio Wkshp] [-]
... Physical examination	PH	the selection and performance of physical examination in the consultation You could have been more thorough in assessing for spread of infection – by checking her temp [Med Report] [-]
... Organisation	OR	overall structure and flow of the interaction; sequence of stages, tasks, actions; process Yeah, I think there was definitely a lack of structure and, sort of preplanning of why he was actually talking to this lady [Physio Wkshop] [-]
<b>Consultation skills</b>	CN	super-ordinate category relating to CS and CL

(continued)

Practitioner skills	PS	super-ordinate category for the following 4 themes
... <b>Professionalism</b>	PF	showing professional responsibility and judgement; preparing for the consultation; managing patients appropriately  I like the way you plan your consultations. Prior to seeing each patient you check their histories, making note of what medications they are taking [Med Report] [+]
... <b>Documentation</b>	DO	1. record-keeping; quality of clinical notes and note-taking 2. use of computer and computer systems; typing skills  Your note taking was perfectly adequate. The main factor that slows you down is your typing [Med Report] [+/-]
... <b>Time management</b>	TM	appropriate organisation and use of time  Questions such as “was there anything else you wanted to discuss today?” can help in time management – as can asking what they are expecting from the consultation or what they hoped you might be able to do for them if this is not clear [reasonably] early on [Med Report]
... .. <b>Prioritisation</b>	PR	ability to set priorities to manage workflow and allocate resources  You handled the seeing of different patients well, while at the same time dealing with an emergency patient with a lacerated face [Med Report] [+]
<b>Knowledge</b>	KN	knowledge-as-resource; underlying knowledge base  But just like I said, that sort of lack of clinical um, aptitude really which she just didn't seem to be able to make those contextual judgments. [Nursing Wkshp] [-]
Language	LA	grammar, intelligibility and accent, oral fluency, pronunciation and intonation, linguistic resources, word choice  I think his questions were at least phrased clearly [Med Wkshp] [+]
Affect	AF	state of mind, mood; affective state; personality trait  at one stage he was clearly getting nervous [Med Wkshp] [-]
World view	WV	beliefs about the world and how people are/ behave, based on the health professional's cultural and social background  P5 would she [doctor] would she even be naïve enough to – or just culturally not aware that – P1 well she might accept the patient's – P5 that this person's just monogamous and that people are just monogamous? (xxx) R1 or she might think it's offensive ((laugh)) P5 offensive yeah [Med Wkshp] [-]

*5.2 A model of what is valued by doctors in the doctor–patient consultation*

The model, developed by the APAI scholar John Pill as part of his doctoral research which focused on the medicine data (Pill, forthcoming), visually represents the aspects relating to a doctor’s performance in the doctor–patient consultation that are valued by participants in this study. The model has three elements, which are presented one above the other in Figure 1; the elements are inter-related, as indicated by the ascending arrow.

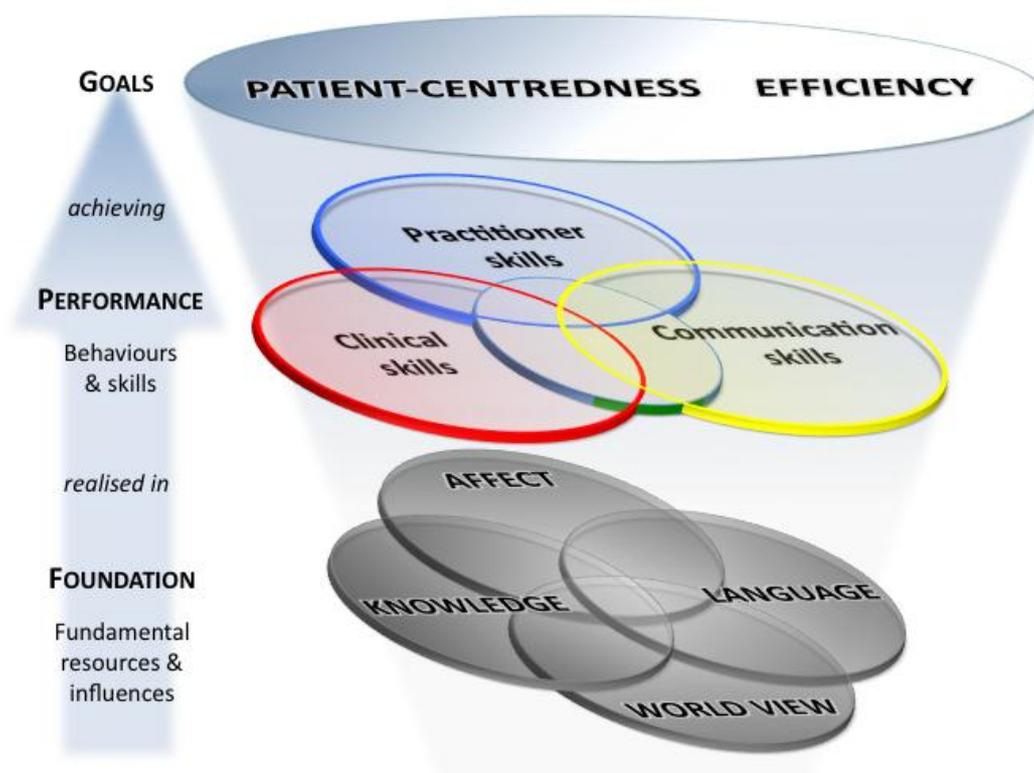


Figure 1. A model of aspects of doctor–patient interaction valued by doctors (from Pill, forthcoming)

The middle element of the model describes the aspects of performance realised in the health professional–patient interaction. The consultation requires the health professional to perform behaviours and skills, which have been categorised here:

*Communication skills*, which involve the health professional’s *Non-verbal communication*, *Manner* and relevant *Interactional tools*. The second set is *Clinical skills*, which relate to *Content* (“knowledge–in–interaction”) and *Physical examination*, *Organisation* and, again, relevant *Interactional tools*. The third and final set is *Practitioner skills*, which concern *Professionalism*, *Documentation*, *Time management*

and *Prioritisation*; relevant *Interactional tools* are also posited as facilitating this set of skills. The skills sets are complementary and the performance of the health professional–patient interaction requires them all, as they are mutually dependent. The components of each set of skills interact with each other and with components of other sets of skills in the performance of the consultation. A repertoire of *Interactional tools* is drawn on as required in the performance of these skills; this is represented by the circle at the centre of this element of the model. These tools include the use of appropriate *Terminology*.

### *5.3 Summary of Phase One*

In Phase One we elicited feedback from three groups of health professionals (doctors, nurses and physiotherapists) on the qualities they saw as important for effective workplace interaction. This feedback was elicited in response to various stimuli, all involving interactions, whether live or simulated, between health professionals and patients. The “indigenous criteria” underlying health professionals’ feedback on these interactions were uncovered via a thematic analysis. The themes so identified formed the basis for a conceptual model (Pill, forthcoming) representing aspects of interaction between health professionals and patients that were valued by health professionals. The task of converting elements of the model to criteria that could be used on a health-specific language test will be described in the following section.

## **6. PHASE TWO: APPLICATION OF PROFESSIONALLY RELEVANT CRITERIA TO OET**

The research question addressed in Phase Two was “Can such professionally relevant criteria be used as the basis for language assessments carried out by *language experts* of migrant health professionals seeking registration in Australia?”

Our task in Phase Two was to determine a) which of the themes identified in the health professionals’ feedback matched the existing OET criteria and b) which additional themes could be considered for inclusion in the OET assessment scheme and which could not. Having established the additional themes to be represented in a revised OET, it was necessary to translate these into assessment criteria which would be meaningful to and useable by language assessors with no prior medical training, to train these assessors to apply these criteria in assessing samples of OET performance and then to

find out how well these criteria functioned both from the assessors' perspective and also in measurement terms. This process and its outcomes will be outlined below.

### *6.1 Themes for inclusion in an expanded OET construct*

While language issues as noted above in Section 5 were seldom mentioned explicitly by the health professionals in our study, there was nevertheless evidence in their feedback that they perceived four of the current OET criteria, namely: INTELLIGIBILITY, FLUENCY, APPROPRIATENESS OF LANGUAGE and RESOURCES OF GRAMMAR AND EXPRESSION as important. (See Pill, forthcoming, for examples of feedback matching these criteria.) Given their importance to health professionals, it seemed appropriate to retain these existing linguistic criteria on OET.

However, a number of themes emerged in the data and associated conceptual model (Pill, forthcoming) that were not reflected in the current criteria. Some of these we considered suitable for inclusion on OET, in the interests of better aligning this test with the health professionals' values. These themes include the full repertoire of *Interactional tools* and *Communication skills* as indicated in the black circles in Figure 2 below.

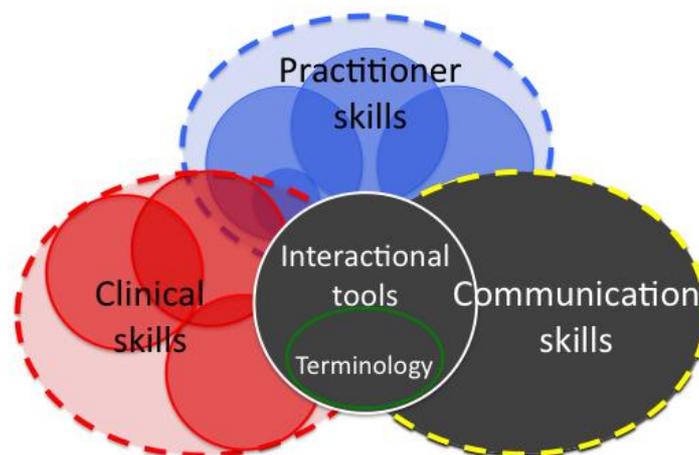


Figure 2. The potential scope of language assessment in the performance element of the consultation  
(from Pill, forthcoming)

Examples of *Interactional tools* from the data set for this study that were deemed to be language-based and therefore recognisable by the OET language assessors are: signposting changes of topic, signposting before asking sensitive personal questions,

structuring the interaction so that it is clear to the patient what is happening, using appropriate information–eliciting techniques, asking open questions and avoiding closed or leading questions, finding out what the patient already knows and what he/she wants to know, and checking if the patient has understood. Behaviours coded under the rubric of *Communication Skills* in the data set which were deemed amenable for inclusion on OET relate to aspects of the health professional’s manner that are realised verbally, such as: supporting the patient’s narrative using active listening, not interrupting the patient, demonstrating a positive, respectful attitude towards the patient, and interacting in an approachable manner.

The other dimensions of the model designated *Clinical skills* and *Practitioner skills* respectively were not considered appropriate for inclusion on a language test. *Clinical skills* (other than the interactional tools mentioned above) involve professional knowledge and therefore cannot be evaluated by language assessors. *Practitioner skills* may also have a clinical knowledge component and require greater opportunities for sustained observation than are available in the OET role–plays. These dimensions were considered outside the boundaries of what could be assessed on a language test.

## *6.2 Operationalising themes as criteria*

There remained the task of transforming the new themes identified above as amenable to assessment on the OET speaking sub–test into a form that would be potentially useable by language assessors without experience in the healthcare workplace. To accommodate the additional themes it was decided to replace the existing criterion OVERALL COMMUNICATIVE EFFECTIVENESS, used on the current OET to make a global judgement of speaking performance, with more precisely defined analytic criteria that would fit better with the professional orientation of the test.

To do so, the instances of each theme considered appropriate for inclusion were organised into sets of similar behaviours. A description of the prototype behaviour was then written for each set, i.e., generalising from the specific instances of behaviour in the data set to an example that described all potential instances of the behaviour. Through an iterative process involving scrutiny of the data extracts and consultation among researchers working on the project, a checklist of 24 descriptions of prototype

## *Towards Improved Healthcare Communication*

behaviours or generalized indicators was produced. These indicators were then grouped into four categories of roughly equal size:

- Indicators of Professional manner (7 descriptions)
- Indicators of Patient awareness (5)
- Indicators for Information–gathering (6)
- Indicators for Information–giving (6)

The first two sets of indicators, Professional manner and Patient awareness, are assumed to be observable in any of the OET role–plays (and by implication in any interaction between health professional and patient in the workplace setting). As for the second two sets relating to Information–gathering and Information–giving respectively, it is recognised that their relevance may depend on the nature of the OET role–play scenarios, some of which place emphasis on the health professional (i.e., the test candidate) giving appropriate information to the patient (the OET interlocutor) while others focus on eliciting information about the patient’s condition. The full checklist of indicators **was provided to the assessors along with a glossary of terms to aid interpretation.**<sup>1</sup>

The 24 indicators in four categories that constitute the checklist were subsequently summarised into two criteria: CLINICIAN ENGAGEMENT derived from the 12 indicators of Professional manner and Patient awareness and MANAGEMENT OF INTERACTION derived from the 12 indicators for Information–gathering and Information–giving. Neither of these aspects of performance has been included in previous assessments of performance on the speaking sub–test of OET. Level descriptors (at four different levels) for these two new criteria were also provided for

---

<sup>1</sup> It is noted that the checklist makes no reference to non–verbal behaviours because these, while deemed by the health professional informants to be important for effective communication, cannot be assessed on the basis of the audio–recordings currently provided to OET assessors. This is a practical limitation of OET, which may need to be reconsidered in future revisions to the test.

rating purposes.<sup>2</sup> The process of training language assessors to apply these new criteria in addition to the existing linguistic ones is described below.

### *6.3 Assessor training*

To help the OET language assessors understand and apply the new criteria, a one-day workshop was held at The OET Centre. The participants at the workshop were the OET assessment manager and assessment officer, seven experienced OET assessors, two workshop leaders and three observers from the project team.

In advance of the workshop the assessment manager had selected 150 previously scored test taker performances from the test data base. This yielded 300 role-play samples (two role-plays for each of 150 candidates). Six of these role-plays were selected for workshop training purposes. The workshop samples included:

- two test takers from each profession investigated in the study (Medicine, Nursing and Physiotherapy);
- a range of score levels across the six performances (from very strong to weak); and
- a mix of gender and first language that reflected the current test taker population.

At the workshop, assessors were briefed about the aims of the whole project and of Phase Two in particular. They were issued with the checklist of indicators and asked to complete the checklist after listening together to one recorded sample. Checklist responses were compared among assessors and discussion ensued about interpretations of particular checklist items. This process was repeated for two more samples. Once they were reasonably familiar with the new criteria, assessors were asked to listen to a further performance, to complete the checklist and to assign ratings against both the old

---

<sup>2</sup> The decision to describe performance at 4 rather than 6 levels (as is the practice with current criteria) was based on a review of the available range of OET speaking performances, which did not lend themselves readily to finer differentiation. It was also hoped that using a different scale would have the added benefit of limiting the “halo effect”, whereby assessors are unduly influenced by the linguistic criteria which they are more familiar with, and thus award the same rating to the new criteria without giving careful consideration to the qualities of performance as described in the checklist.

(linguistic) criteria and the new ones. Checklist responses and ratings were again discussed and the process was repeated for a further two OET speaking samples.

Following the workshop the assessors listened to and rated 80–90 further performances (pseudo–randomly ordered) using the checklist and criteria (new and existing). They were asked to use the checklist to help them orient to the criteria for the first several performances (or as many as they felt necessary); they were also asked to “refresh” this orientation by using the checklist for a while when resuming rating after a break. The distribution of samples was organised to ensure that each performance was rated by two or more different assessors.

#### *6.4 Functioning of new criteria*

The functioning of new criteria from the perspective of the participant assessors was determined by analysing their feedback both on the workshop and on the subsequent rating process. Assessors were asked to provide feedback on their confidence in a) understanding the meaning of the two new criteria, b) using the checklist to inform their ratings and c) applying the level descriptors for each of the criteria.

Findings, summarised in Table 4 below, suggest a satisfactory degree of confidence in understanding the new criteria, using the checklist and applying the level descriptors, with mean scores for each item ranging between 3.6 and 4 on the five–point Likert scale. Open–ended comments from assessors also revealed considerable enthusiasm for the new initiative. They felt that the additional criteria offered more explicit guidance than the rather vague OVERALL COMMUNICATIVE EFFECTIVENESS category that they replaced. The new criteria were also viewed as encouraging a focus on the patient and as acknowledging interactive skills rather than simply focussing on an individual’s language competence.

Table 4. Self–reported rater confidence; 1 (not all) to 5 (extremely)

	<b>Clinician Engagement</b>	<b>Management of Interaction</b>
	Mean (Range)	Mean (Range)
Recognising aspects of criterion	3.9 (3–5)	3.7 (3–4)
Applying checklist	4.0 (4–4)	3.7 (3–4)

The functioning of these ratings in measurement terms was determined through a number of statistical analyses. The first analysis, using the multidimensional Rasch program Facets (Linacre, 2008), tested the proposition that the different sets of criteria (old and new) were measuring similar constructs of ability. Two individual Facets runs were undertaken: one using scores assigned by assessors on the new criteria and the other using scores assigned on the existing language criteria. The two sets of candidate measures (in logits) yielded by these separate analyses were then correlated. The resultant correlation of 0.82, represented graphically in the scatterplot below (Figure 3), suggests that the two sets of criteria are ranking the candidates with the degree of similarity we would expect from a typical pair of judges rating the candidates. The tentative conclusion to be drawn from this analysis is that ratings from the two sets of criteria could be combined to produce a single score.

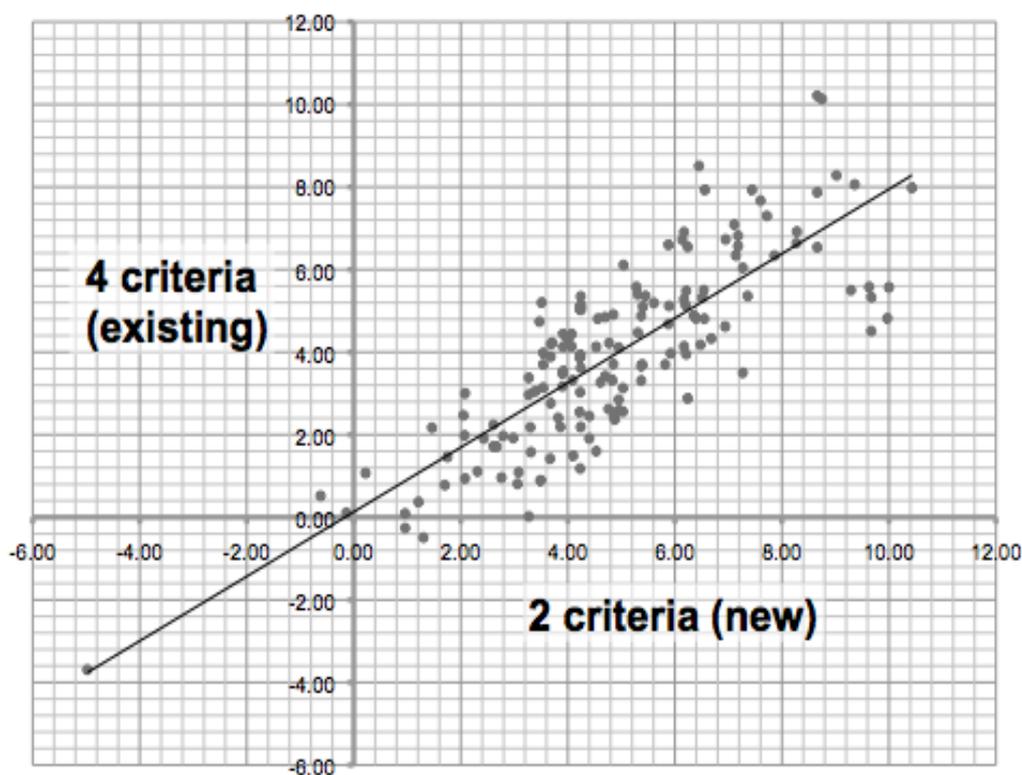


Figure 3. Relationship between scores (in logits) for existing and new criteria

This conclusion was tested via a further Facets analysis in which scores on all criteria (existing and new) were entered into the model. This purpose of this analysis was to test the unidimensionality assumption of the Rasch measurement model or, in other words,

the notion that both sets of criteria were collectively measuring a single underlying construct of ability. Results showed a very high level of model fit explaining 75.3% of the variance in scores. Again this supported what was suggested by the previous correlational analysis, namely that the current and new criteria are measuring a common construct of ability.

A more rigorous test of the unidimensionality assumption was undertaken in the form of a Principal Component Analysis of residuals using the Winsteps program (Linacre, 2013). This was to determine whether score variance that was not explained by the single model assumed in the Rasch model was attributable simply to random error or instead to the presence of other constructs in the mix. Results presented in Figure 4 below show that the strongest “other” measure, apart from the primary language construct, is related to the two new criteria (numbered one and two in the top left quadrant). While we must be cautious in our interpretation of this finding given that the strength of the contrast (1.7 eigenvalues) is slightly below the conventionally required threshold (of 2 eigenvalues), it seems reasonable to suggest that the new criteria, while overlapping substantially with the existing ones, are also adding a new dimension to the measurement model, or in other words extending the scope of the test construct. Based on this interpretation one could argue for separate reporting of scores on the existing and new criteria to provide a more nuanced picture of candidate ability.

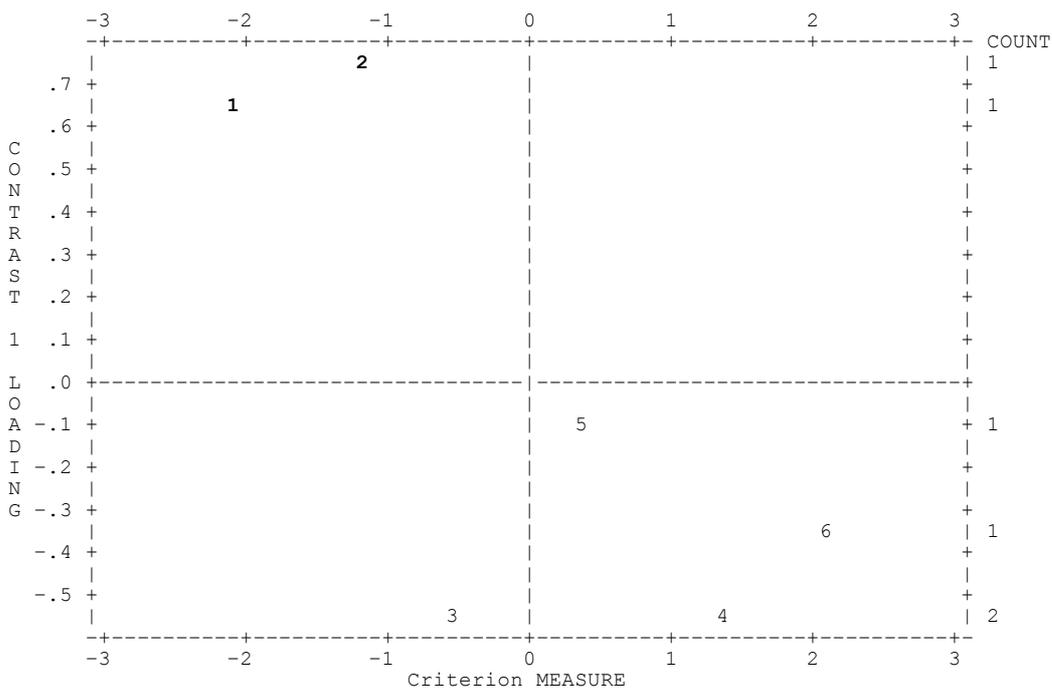


Figure 4. Plot of standardised residual contrast

Further data analysis and elaborated discussion of these issues will be presented in publications relating to this second phase of the study. This will help to form a stronger basis for deciding on the practical question of whether to report a single score or separate scores, a decision for which The OET Centre is ultimately responsible.

### *6.5 Summary of Phase Two*

The answer to the research question posed in relation to this phase of the study is that yes, on the basis of the evidence presented above, it is possible for language assessors to use professionally–relevant criteria in rating performance on OET through the provision of a checklist which assists them in identifying the relevant features of performance and appropriate training in applying the checklist and criteria. Feedback from a sample of OET language assessors, following a training session and subsequent rating practice, indicated confidence in substituting the new professionally relevant criteria, CLINICIAN ENGAGEMENT and MANAGEMENT OF INTERACTION, for the holistic OVERALL COMMUNICATIVE EFFECTIVENESS criterion and using these alongside the linguistic criteria currently used for assessing performance on OET, namely: INTELLIGIBILITY, FLUENCY, APPROPRIATENESS OF LANGUAGE and RESOURCES OF GRAMMAR AND EXPRESSION. Statistical analyses suggest that the new criteria are psychometrically compatible with the current linguistic criteria, while at the same time adding something new to the assessment, as was our intention in undertaking the project.

## **7. PHASE THREE: SETTING MINIMUM STANDARDS FOR PROFESSIONAL REGISTRATION**

The Research Question addressed in Phase Three, “What *minimum standards* in terms of these criteria should be set for professional registration of migrant health professionals?” involved a standard setting exercise for the OET speaking test using the judgements of clinical educators in medicine, nursing and physiotherapy on selected audio–recorded role–play performances across a range of proficiency levels. In this way, three sets of views of the performance of candidates could be compared.

From Phase Two, two sets of scores for each candidate were available: existing scores from recordings which had been previously rated against the existing criteria; and new

scores from the Phase Two workshops which were scored against the new, expanded set of criteria. The aim of Phase Three was to see how the judgements of clinical supervisors and educators lined up with these two sets of scores. Two issues were at stake, each involving potential differences between the views of each set of judges: the OET assessors in the Phase Two workshop using the new criteria, and clinical educators and supervisors. First, was there a difference in the standards applied in the current operational version, using the existing criteria, and the standards applied by the health professionals? That is, would the same individuals be classified as belonging to the same categories? Would the same candidates pass? Second, was there a difference in the standards applied by the OET assessors using the new criteria, and the standards applied by the health professionals? Would candidates currently passing, when scored against existing criteria, still be likely to pass if their performances were scored against the new criteria?

### *7.1 Data elicitation*

In order to address these questions, a number of workshops for clinical supervisors/educators were held, separately for each of medicine, nursing and physiotherapy. The aim of the workshops was to elicit judgements from the health professionals as to the readiness of the candidate, on the evidence available from the recorded role-play performances, to cope with spoken communicative tasks in the clinical setting. A rating sheet was prepared with seven categories. Four were anchor categories (“Strong”, “Competent”, “Not Yet Competent” and “Unsatisfactory”), with three intermediate points between these (“Between Strong and Competent”, “Between Competent and Not Yet Competent”, “Between Not Yet Competent and Unsatisfactory”). At least three role-play performances were played and workshop participants scored them independently, and then discussed their reasons. The participants were then given 20 further digital recordings to assess following the workshop, and these assessments were the data used in the subsequent analysis. In medicine, 13 health professional judges rated each of 23 candidate performances; in nursing, 18 judges rated each of 26 performances; in physiotherapy, 8 judges rated each of 24 performances.

## *7.2 Findings*

The analysis took two forms: use of the analytic judgement method (Plake & Hambleton, 2001) for setting cut-scores, and multi-faceted Rasch measurement (implemented in the program Facets [Linacre, 2008]) of the judgements.

In the analytic judgement method, for each profession, each candidate was examined to see how they had been classified by the judges, and this was compared with the score they had got from the OET assessors. A simple arithmetic procedure was used to calculate the average score for each judgement level, and this was set as the cut-off for that level. The cut-scores differed slightly for each profession; the cut-scores for nursing were lowest, which raises the possibility that different cut-scores should be used for each profession in operational administrations. There was one small inconsistency in the cut-score level for physiotherapy at the lower levels, which was caused by the paucity of candidates at these levels; more data would be needed to resolve this issue.

In this way it was possible to compare the classification of individuals (as grades A, B, C or D) using the existing cut-offs (and based on the old criteria) with the classification using the new cut-offs (and based on the new criteria). It showed that using the new cut-scores, and the new criteria, would have an impact on rates of passing overall, and of classification into the different categories. For example, for medicine, more candidates would be given A (moved up from B to A), but fewer candidates overall would pass (there would be fewer Bs and more Cs). For nursing, the impact on overall pass rates, although they would be lower, would be less, but with many more nurses classified as A rather than B. For physiotherapy, there would be no impact on overall pass rates, but again with more candidates classified as A rather than B.

In the Facets analysis the scores from the health professionals on the 7-point scale were analysed. This was done to examine the consistency and coherence of judgement behaviour within and among the cut-score judges. It also yielded an overall measure for each candidate, which could be compared with the measures yielded from the scores of the regular OET raters using the new, expanded set of criteria.

The results for medicine showed that all the judges were rating consistently, though with somewhat different standards of harshness and leniency in terms of allocation to

the given categories, an entirely expected result, and one in line with previous studies of rating behaviour. Two of the 23 performances rated were found to be challenging for the judges to agree on. The reliability of the candidate measures derived from the ratings was 0.92; this was an acceptable result. When the measures resulting from the health professional judgements were compared with the measures resulting from the OET rater judgements were compared, there was good agreement in the ranking of candidates; all but 4 of the 23 candidates were ranked in the same order. For nursing, there was slightly lower level of agreement: 6 of 26 candidates were ranked in a different order, although most disagreement was in the C and D range, which has no consequences for pass/fail decisions. For physiotherapy, 3 of the 24 performances were ranked in a different order, but again this was among the weaker candidates.

More detailed analysis of the available data will appear in manuscripts for publication on Phase Three.

### *7.3 Summary of Phase Three*

Overall, this phase was successful in establishing new cut-scores, and suggested that the impact of the new criteria would be significant in the case of medicine in particular. The inclusion of the new criteria for many candidates across the three professions confirmed more strongly that they had the necessary skills for successful communication in clinical settings; for a minority of candidates, the inclusion of the new criteria suggested greater caution in accepting them as ready for clinical placement.

## **8. RECOMMENDATIONS FOR PRACTICE**

Recommendations for practice based on the study findings can be divided into three major areas. These areas include: changes to OET, training requirements, consultation with stakeholders and dissemination of outcomes.

### *8.1 Changes to OET*

In terms of proposed changes to OET, it is suggested that two new professionally relevant criteria be included in addition to the existing linguistic criteria. These new criteria, which have been derived from a careful analysis of health professional feedback gathered in Phase One on actual or simulated communication in healthcare

settings, are CLINICIAN ENGAGEMENT and MANAGEMENT OF INTERACTION. According to the OET assessors involved in trialling these criteria, they offer more explicit information on the overall qualities of candidate performance and can therefore replace the current holistic criterion OVERALL COMMUNICATIVE EFFECTIVENESS. OET assessor feedback also indicated that the new criteria encouraged a focus on patient interaction with the clinician and on interactive skills rather than a focus on just language competence. Furthermore, our statistical analyses from a pilot study with OET assessors (Phase Two) suggest that, from a measurement perspective, these new criteria are compatible with the existing ones and yield sound estimates of candidate ability.

The addition of the new criteria has implications for the way candidates' results are reported. Data from Phase Two of the study suggest two possible options: a combined score which subsumes scores on both the current linguistic criteria and the new criteria or separate scores on the existing and new criteria to allow a more nuanced picture of candidate ability. Factors to consider in selecting between these options are the benefits (greater diagnostic power) and costs (e.g. in time) for OET in processing and delivering results and for OET users in interpreting and using the additional information provided.

The OET Centre also needs to examine the ways in which health professional test takers are classified. As described in Phase Three, inclusion of new cut-off scores using the new criteria will potentially lead to greater confidence of the attributes held by candidates, since these cut-offs are based on the communicative qualities which health professionals consider to be important and on a consensus view about what minimum standards are required. Decisions about whether to apply these new cut-scores should be made with awareness of their potential impact on pass rates for each of the targeted professions and on the nine other professions using OET.

## *8.2 Training requirements*

In terms of training requirements, if the new criteria are adopted, additional support will be needed for guiding OET language assessors in the conduct of their activities. In recommending the inclusion of two additional criteria in OET, the checklist developed for this study (Appendix 3) should be used to help in training OET assessors in understanding the new criteria and in applying the level descriptors. To learn about the application of the new criteria, a workshop can be conducted with OET assessors. The

length of the workshop should be determined by the learning needs of assessors. In the study, a one-day workshop was found to be effective and adequate in supporting a small group of OET assessors to develop understanding and confidence in using the additional criteria and to produce acceptably consistent rating behaviour. Ongoing monitoring of assessor behaviour should of course continue after the initial training session, in line with current practice at The OET Centre.

### *8.3 Consultation and dissemination*

Three professional groups were used in the conduct of this study, medicine, nursing, and physiotherapy, yet OET is used to determine minimum English communication standards for twelve health professional groups. While the research team believes it is feasible and appropriate for the two new criteria to be taken up for OET by all twelve professional groups, consultation regarding their suitability and uptake is needed with expert representatives from each profession. In addition, consultation is required with the health professional boards to consider how any such changes to OET can be implemented and disseminated. These consultations should be undertaken not only with the three health professional groups represented in this study, but also with the nine health professional groups that were not involved.

Any changes to OET arising from the above consultations will need to be disseminated not only to the health profession at large to assist them in understanding and interpreting OET results, but also to test takers and their teachers, since the introduction of the new criteria will certainly influence how candidates prepare or are prepared for the test. Making the checklist available to test users will potentially increase confidence in the test's relevance to the healthcare context and acculturate test takers to the expectations of the Australian workplace.

## **9. RECOMMENDATIONS FOR FURTHER RESEARCH**

This study has drawn attention to a number of research avenues worth pursuing in the interests of consolidating and enhancing the validity and public standing of OET as a specific purpose test of communication in healthcare settings.

### *9.1 Validating the new criteria with additional professions*

While it has been suggested above that the new criteria can be applied to assessing performance of role–plays across the full range of professions served by OET, it would be appropriate to confirm this empirically by trialling the criteria on a wider sample of existing role–play performances. This could be undertaken as a follow–up to training the entire cohort of OET assessors in the application of the new criteria as recommended in Section 8.2 above.

### *9.2 Setting cut–scores with additional professions*

New cut–scores have been proposed for the three health professions included in the current study. Following the model adopted for this project, additional standard setting workshops should be conducted with the remaining health professions and the impact of resultant changes to cut–scores on pass rates should be considered.

### *9.3 Exploring the role of non–verbal behaviour*

Non–verbal behaviours were regarded by health professional informants in Phase One of this study to be important for effective communication, but have thus far been excluded from consideration on OET. It was not appropriate to include non–verbal behaviours on the checklist developed in Phase Two, given that, for practical reasons, the speaking test is assessed using audio–recordings only. It is recommended that further research be undertaken using video–recorded OET role–plays and incorporating non–verbal behaviours on the assessor checklist to determine their influence on assessment outcomes, compared to the other listed aspects of CLINICIAN ENGAGEMENT and MANAGEMENT OF INTERACTION. Such a study could include a comparative dimension, whereby the same role–play performances were assessed in both audio– and video–recorded formats. Findings would indicate the cost of excluding non–verbal behaviours from the assessment of communicative effectiveness and thereby inform policy regarding any future revisions to the OET speaking test.

### *9.4 Understanding a wider range of interactions in healthcare*

OET is designed to assess the effectiveness of spoken communication based on role–plays between health professionals and patients. While interaction between health

professionals and patients is critical for ensuring patient safety and optimal health outcomes and is rightfully prioritised on OET, much of patient care is conducted inter- and intra-professionally, on the telephone, in hand-over meetings, multi-disciplinary meetings, and at the bedside of patients. Inter- and intra-professional spoken communication is a key component of healthcare communication and, in the interests of better construct representation, may warrant inclusion in a language test designed for health professionals. Before considering this, a better understanding of the language and communication skills involved in these additional areas of healthcare communication is needed, including the extent to which the discourse demands of these interactions differ from those between health professionals and patients that are currently represented on OET. Observation and discourse analysis of a broader range of workplace interactions are therefore recommended for future investigation.

#### *9.5 Written communication in healthcare settings*

The emphasis on this study has been on spoken communication in establishing relevant and rigorous English communication standards for how non-native English speaking health professionals interact with patients. The research team has recently obtained funding for a second ARC Linkage Project grant (LP130100171) to examine written practices of communication between non-native English speaking health professionals and their native English speaking colleagues. Following conduct of this project, it will be possible to provide comprehensive recommendations about written communication and about interactions between spoken and written communication in the healthcare settings. These recommendations will be used in developing spoken and written English communication standards to be applied to non-native English speaking health professionals seeking to practise in Australia.

## REFERENCES

- Australian Government Department of Health and Ageing. (2008, April). *Report on the audit of health workforce in rural and regional Australia*. Canberra: Commonwealth of Australia.
- Barton, D., Hawthorne, L., Singh, B. & Little, J. (2003). Victoria's dependence on overseas-trained doctors in psychiatry. *People and Place*, 11(1), 54–64.
- Basturkmen, H. & Elder C. (2004). The practice of LSP. In A. Davies & C. Elder, (Eds.), *Handbook of applied linguistics* (pp. 672–694). Oxford: Blackwell.
- Birrell, B., Rapson, V. & Smith, T. F. (2005). *Immigration in a time of domestic skilled shortages: Skilled movements in 2003–04. A report to the Department of Immigration, Multicultural and Indigenous Affairs (DIMIA)*. Canberra: Commonwealth of Australia.
- Birrell, B. & Schwartz, A. (2006). Accreditation of overseas trained doctors: The continuing crisis. *People and Place*, 14(3), 1–11.
- Brown, A. (1995). The effect of rater variables in the development of an occupation-specific language performance test. *Language Testing*, 12(1), 1–15.
- Cizek, G. J. & Bunch, M. B. (2007). *Standard setting: A guide to establish and evaluating performance standards on tests*. London: Sage Publications.
- Cizek, G. J., Bunch, M. B. & Koons, H. (2004). Setting performance standards: Contemporary methods. *Educational Measurement*, 23(4), 31–49.
- Dias, P., Freedman, A., Medway, P. & Paré, A. (1999). *Worlds apart: Acting and writing in academic and workplace contexts*. Mahwah, NJ: Erlbaum.
- Douglas, D. (2000). *Assessing languages for specific purposes*. Cambridge: Cambridge University Press.
- Douglas, D. (2001). Language for specific purposes assessment criteria: Where do they come from? *Language Testing*, 18(2), 171–186.

- Douglas, D. & Myers, R. (2000). Assessing the communication skills of veterinary students: Whose criteria? In A. J. Kunnan, (Ed.), *Fairness and validation in language assessment: Selected papers from the 19<sup>th</sup> Language Testing Research Colloquium, Orlando, Florida* (pp. 60–81). Cambridge: Cambridge University Press.
- Douglas, S. (2008). The registration and accreditation of international medical graduates in Australia – a broken system or a work in progress? *People and Place*, 16(2), 28–40.
- Eggly, S., Musial, J. & Smulowitz, J. (1999). The relationship between English language proficiency and success as a medical resident, *English for Specific Purposes*, 18(2), 201 – 208.
- Elder, C. (1993). How do subject specialists construe classroom language proficiency? *Language Testing*, 10(3), 235–254.
- Elder, C. (2001). Assessing the language proficiency of teachers: Are there any border controls? *Language Testing*, 18(2), 149–170.
- Elder, C., Pill, J., Woodward–Kron, R., McNamara, T., Manias, E., Webb, G. & McColl, G. (2012). Health professionals' views of communication: implications for assessing performance on a health–specific English language test. *TESOL Quarterly*, 6(2), 409–419.
- Hawthorne, L. (2012). International medical migration: What is the future for Australia? *Medical Journal of Australia Open*, 1(Suppl. 3), 18–21.
- Hawthorne, L. & Birrell, B. (2002). Doctor shortages and their impact on the quality of medical care in Australia. *People & Place*, 10(3). 55–68.
- House of Representatives Standing Committee on Health and Ageing. (2012). *Lost in the labyrinth: Report on the inquiry into registration processes and support for overseas trained doctors*. Canberra: Commonwealth of Australia.

- Jacoby, S. W. (1998). *Science as performance: Socializing scientific discourse through conference talk rehearsals*. Unpublished PhD thesis, University of California, Los Angeles.
- Jacoby, S. & McNamara, T. (1999). Locating competence. *English for Specific Purposes*, 18(3), 213–241.
- Linacre, J. M. (2008). *Facets Rasch model computer program* [Software]. Chicago: Winsteps.com.
- Linacre, J. M. (2013). *Winsteps Rasch measurement computer program* [Software]. Beaverton, Oregon: Winsteps.com.
- McDonnell, L. & Usherwood, T. (2008). International medical graduates: Challenges faced in the Australian training program. *Australian Family Physician*, 37(6), 481–484.
- McNamara, T. F. (1996). *Measuring second language performance*. London: Longman.
- O'Hagan, S., Manias, E., Elder, C., Pill, J., Woodward–Kron, R., McNamara, T., McColl, G. & Webb, G. (in press). What counts as effective communication in nursing? Evidence from nurse educators' and clinicians' feedback on nurse interactions with simulated patients. *Journal of Advanced Nursing*.
- Pill, T. J. H. (forthcoming). *What doctors value in consultations and the implications for specific-purpose language testing*. PhD thesis, University of Melbourne.
- Plake, B. S. & Hambleton, R. K. (2001). The analytic judgement method for setting standards on complex performance assessments. In G. J. Cizek, (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp. 283–312). Mahwah, NJ: Erlbaum.
- Woodward–Kron, R., van Die, D., Webb, G., Pill, J., Elder, C., McNamara, T., Manias, E. & McColl, G. (2012). Perspectives from physiotherapy supervisors on student–patient communication. *International Journal of Medical Education*, 3, 166–174.

### ***Towards Improved Healthcare Communication***

Xu, Y. (2010). Transitioning international nurses: An outlined evidence-based program for acute care settings. *Policy, Politics & Nursing Practice*, 11(3), 202–213.

Zieky, M. J. (2001). So much has changed: How the setting of cut scores has evolved since the 1980s. In Cizek, G. J., (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp. 19–51). Mahwah, NJ: Erlbaum.

## **APPENDIX 1: PROJECT TEAM**

### **Chief investigators**

- Assoc. Prof. Catherine Elder, formerly Director Language Testing Research Centre; Principal Fellow, School of Languages and Linguistics, University of Melbourne.
- Prof. Tim McNamara, Professor in Linguistics and Applied Linguistics, School of Languages and Linguistics, University of Melbourne.
- Prof. Elizabeth Manias, Professor of Nursing, School of Health Sciences, University of Melbourne.
- Prof. Geoff McColl, Director, Medical Education Unit, Melbourne Medical School; Deputy Dean, Faculty of Medicine, Dentistry and Health Sciences, University of Melbourne.
- Assoc. Prof. Gillian Webb, formerly Assoc. Prof. of Physiotherapy; Head, School of Physiotherapy, University of Melbourne.
- Assoc. Prof. Robyn Woodward–Kron, Assoc. Prof. in Healthcare Communication, Medical Education Unit, Melbourne Medical School, University of Melbourne.

### **Australian Postgraduate Award (Industry) Scholar**

- Mr John Pill, Language Testing Research Centre, University of Melbourne

### **Other team members**

- Dr Sally O’Hagan, Project Officer, Language Testing Research Centre, University of Melbourne.
- Dr Diana van Die, Freelance Research Assistant.

**APPENDIX 2: REFERENCE GROUP MEMBERS**

- Ms Fel Bisiani, Manager, The OET Centre.
- Dr Ying Zhang, Assessment Team Leader, The OET Centre.
- Dr Joan Deegan, Lecturer, Division of Nursing and Midwifery, La Trobe University.
- Dr Sean Fabri, Medical Clinical Educator, Western Health; IMG Subcommittee Chair, Postgraduate Medical Council of Victoria.
- Dr Eleanor Flynn, Senior Lecturer in Medical Education, Medical Education Unit, University of Melbourne.
- Ms Janine Loader, Chief Nursing Officer, Director of Nursing, Mercy Hospital.
- Prof. Joan McMeeken, Professor of Physiotherapy, University of Melbourne.
- Ms Cathy Nall, Director of Physiotherapy, Austin Health.