Exploring Culturally Sensitive Alternatives for the First Certificate in English

E. J. O'Donovan
Acknowledgements

The author gratefully acknowledges the advice and encouragement of dissertation supervisor Dr Hitomi Masuhara, Dr Lesley Randles and that of Liverpool University English Department’s academic and support staff. Sincere thanks also go out to the research participants, without whom this paper would not be possible. The contributions of all reviewers, translator Mukhtar Ellakhmi and marking consultant Aleksandra S. Mrazek are also greatly appreciated.

Dedicated to my family and the memory of my grandfather,
Dr Richard Henry Birch  MDS  DDO  FDSRCS.Ed  FRSH  BDS JP (1927-2014)
Abstract

Throughout recent years, Cambridge English have expressed a steadfast commitment to fairness in the language testing arena. However, the cultural dimension to this commitment is often left under-defined or subsumed under broader terminology. Despite its commissioning of detailed research in the area, the cultural suitability of the visuals used in Cambridge's actual examination content remains an area of some ambiguity.

Informed by the views and performances of a small but representative sample of young Kuwaiti nationals, this pilot study specifically focuses on the cultural orientation of the photographic representations used in Cambridge English's First Certificate in English examination [FCE]. Finding that many such representations do exhibit strong Anglo/American centricity, it seeks to explore the use of more personal and accessible testing alternatives. These alternatives were created with learner authentication and mental representation opportunities in mind, in the hope that deeper sensory engagement and personalisation would minimise the need for explicitly cultural content. In essence, it sought to reorientate the examination around that which all candidates share; namely their senses and past experience. Despite some methodological caveats, sharp increases in output and affective engagement were recorded throughout the pilot, giving every indication of a promising avenue of further enquiry.

It is argued that, if Cambridge English are to be truly consistent on grounds of fairness, they need to reconsider the cultural content of their photographic representations when test-writing for such a diverse global audience. The current pictorial format may not meet the standards of neutrality Cambridge English set themselves and alternatives, whether those suggested in this paper or otherwise, demand a closer look.
## Contents

**Chapter 1: Introduction:**
- 1.1 Motivation ................................................................. 5
- 1.2 Statement of Purpose .................................................... 6
- 1.3 Need for Research ................................------------------- 5
- 1.4 Structure of this Study ................................................. 7

**Chapter 2: Key Concepts & Literature Review:**
- 2.1 Cambridge Language Testing & the FCE .......................... 7
- 2.2 Culture & Cultural Validity ........................................... 9
  - 2.2.1 Whose Culture? Whose English? ............................... 10
- 2.3 Cambridge’s Commitment to Fairness .............................. 11
  - 2.3.1 Overview of FCE Claims ......................................... 11
- 2.4 Proposed Solutions ..................................................... 13
  - 2.4.1 Learner Authentication ........................................... 14
  - 2.4.2 Multi-Dimensional Mental Representation .................... 15
    - 2.4.2.1 Experiential Learning ........................................ 17
    - 2.4.2.2 Neurological & Psychological Background ................ 18
    - 2.4.2.3 MDMR & the ‘Multi-Dimensional Approach’ ............. 20

**Chapter 3: Methodology:**
- 3.1 Overview and Rationale of this Research ......................... 22
- 3.2 Structure of the Speaking Paper .................................... 23
- 3.3 Participants ................................................................ 24
- 3.4 Procedure .................................................................. 24
- 3.5 Research Limitations and Concerns ................................ 27

**Chapter 4: Analysis:**
- 4.3 Background Perspectives .............................................. 29
  - 4.3.1 Candidate Comments ............................................ 30
    - 4.3.1.1 Initial Comments ............................................. 30
    - 4.3.1.2 Specific Comments ........................................... 32
- 4.4 Suggested Alternatives ................................................ 34
  - 4.4.1 Alternative Item 1: ‘Audio Response’ ......................... 35
  - 4.4.2 Alternative Item 2 (i): ‘Silhouette’ ............................. 36
    (ii): ‘Picture Table’ .................................................... 37
  - 4.4.3 Alternative Item 3: (i) ‘Eyes-Closed’ [Solo] .................. 38
    (ii) ‘Eyes-Closed’ [Collaborative] .................................. 39
  - 4.4.4 Alternative Item 4: ‘Personal Picture Description’ ......... 40
- 4.5 FCE Item 1: ‘Museums’ ................................................. 41
  - 4.5.1 Transcript ............................................................ 42
  - Observations:  4.5.2 Content Engagement .......................... 42
    - 4.5.3 Grammar & Vocabulary ....................................... 43
    - 4.5.4 Output, Pronunciation & Fluency ............................. 43
    - 4.5.5 Marks ............................................................... 43
- 4.6 ‘Eyes-Closed’ [Solo] Alternative Item ............................... 44
  - 4.6.1 Transcript ............................................................ 44
  - Observations:  4.6.2 Content Engagement .......................... 44
    - 4.6.3 Grammar & Vocabulary ....................................... 44
    - 4.6.4 Output, Pronunciation & Fluency ............................. 45
    - 4.6.5 Marks ............................................................... 46
- 4.7 FCE Item 2: ‘School Activities’ ...................................... 46
  - 4.7.1 Transcript ............................................................ 47
  - Observations:  4.7.2 Content Engagement .......................... 47
    - 4.7.3 Grammar & Vocabulary ....................................... 48
    - 4.7.4 Output, Pronunciation & Fluency ............................. 48
    - 4.7.5 Marks ............................................................... 49
- 4.8 ‘Personal Picture Description’ Alternative Item .................. 49
  - 4.8.1 Transcript ............................................................ 50
  - Observations:  4.8.2 Content Engagement .......................... 51
    - 4.8.3 Grammar & Vocabulary ....................................... 51
    - 4.8.4 Output, Pronunciation & Fluency ............................. 52
    - 4.8.5 Marks ............................................................... 52

**Chapter 5: Conclusion** ................................................... 53
**Bibliography** ................................................................... 58
**Appendix** ........................................................................ 62
Chapter 1: Introduction

1.1 Motivation

My prior professional and academic experience is firmly rooted in the International Development sector, specifically in communications. The extent to which certain constructs, practices and modes of expression carry with them cultural significance has long been of interest to me. This interest was thrown into sharp perspective two years ago when I began teaching on the outskirts of Saudi Arabia’s most conservative major city, Riyadh. I was working in a ‘communicative’ conversational capacity with exclusively male learners, some of whom were preparing for Cambridge Assessment. All students were Middle-Eastern, though within that context their nationalities were diverse (Saudi, Yemeni, Kuwaiti, Syrian). Most had extremely limited direct exposure to Western culture.

The Kingdom of Saudi Arabia (KSA) has undergone some significant shifts over the past twenty years and is currently spending more on education per capita than any other nation globally. A total expenditure of £37 bil is projected for 2015, representing 25% of the government’s entire budget (Foreign and Commonwealth Office: 2015). Many mainstream domestic university programs are now taught in English and this development has encouraged widespread enrolment in private ELT institutes. Whereas some improvements are evident (Wiseman, Sadaawi and Alromi: 2008), the KSA continues to perform very poorly when judged against common international standards (with an IELTS mean of 5.0 and a FCE failure rate fluctuating between 40% and 70% in recent years) (Cambridge English 2015e).

The reasons for this underperformance are undoubtedly diverse and have been the subject of some debate (Wiseman, Sadaawi and Alromi: 2008). Low motivation, excessive educational bureaucracy, high teacher turn-over and narrowness of curriculum have all been proposed as inhibitors to educational progress (Alamri: 2011, Liton: 2012, Anthony: 2014). While teaching in the KSA, I slowly began to view the content of formal assessment as being a contributing factor, one detrimental to the testing and learning experience of students who are culturally distant from what we might term the ‘western experience’. This seemed particularly evident in B2-level English language speaking papers in which students often struggled to relate to, or even identify, the visuals chosen for discussion. I came to the conclusion that needlessly culturally-engendered images were preferred to more open spaces for personal self-expression. It struck me that a more personal, experiential approach could offer a means of circumventing these cultural problems by re-centring assessment around the students themselves. I wish to take the opportunity presented by this dissertation to investigate such potential.
1.2 Statement of Purpose

The initial purpose of this project is to look at the way one such examination (Cambridge's First Certificate in English [FCE]) uses imagery as a basis for discussion in an examination setting. This will be done to establish whether any evidence exists for cultural bias and exclusion. If found in evidence, it is hoped that the accessibility of these test items can be improved through personalisation and opportunities for multi-dimensional mental representation (MDMR). Loosely defined, MDMR is the means by which we internally visualise and process both the world around us and our past experiences (see chapter 2, p.15 for a detailed description). Appeals to it have been linked with the higher cognitive and affective engagement associated with durable language acquisition (Masuhara: 1998, Tomlinson: 2000). The main hypothesis driving this paper is that MDMR can potentially add a new dimension to language testing and could help ameliorate cultural bias found therein. Rather than ignoring cultural bias or making materials so anodyne that they are stripped of their affective qualities, it is hoped that students can be offered the flexibility to generate, reflect on and thereby personalise their own test content in a 'learner-authentic' way. The ultimate goal of this piece is to determine what effect this more open approach to testing might have on the performance of young adults who fit the above profile.

1.3 The Need for Research

There has been very little published research into the prospect of utilising MDMR for testing purposes and, thus far, no substantive efforts have been made to apply such techniques to mainstream international assessment. As such, no research has been conducted into the prospect of countering international test bias by appealing to that which we all share – our senses. The potential of MDMR is often alluded to favourably in the academic literature surrounding the field (Arnold: 1999, Tomlinson: 2000, Masuhara: 2005), but its application has typically been focussed on classroom teaching and materials design.

Unlike coursebooks however, international examinations operate as closed systems. They cannot be adapted by and filtered through an intermediary teacher for a specific audience. The learner must receive and interact with them directly in their unaltered state. Any cultural or contextual problems found therein therefore proceed uninhibited. This makes the area of improving any deficiencies in international assessment particularly important, one with clear and direct consequences for the learner. Teachers, learners and examiners alike have a vested interest in efforts to improve the fairness of international assessment wherever weaknesses are evident. Whether the potential of MDMR can translate from the classroom to the examination room is therefore a viable avenue of enquiry.
1.4 Structure of this Study

Having outlined the motivation behind, purpose of and justification for this research, we now turn to how the remainder of this paper is structured. The relevant academic contributions underpinning it (specifically in the areas of Cambridge testing, cultural bias, learner-authentication and MDMR) will be covered in the literature review below. The methodology for identifying and improving deficiencies in the examination will be outlined in chapter three (p.22). Accompanying this will be a 'research limitations' section - a critical look at this paper's own limitations and weaknesses, necessary to qualify the subsequent investigation (p.27). An evaluation of the FCE's photographic representations, driven by the observations of the candidates themselves, will follow (pp.29-34). Designed with the provision of MDMR opportunities in mind, alternative suggestions for test items will then be laid out and critically evaluated to determine which are suitable for implementation in this context (pp.35-40). Analysis of participant performance will follow, assessed against Cambridge's own speaking paper rubrics (pp.41-52). The remaining chapter will be devoted to conclusions, implications and recommendations for Cambridge's ongoing efforts in the testing field (pp.53-57).

Chapter 2: Key Concepts and Literature Review

This project will necessarily draw upon a wide range of literature from a number of academic disciplines. The areas focused on in this review broadly fall into the categories of the First Certificate in English examination itself, Cultural Bias, Learner Authentication and MDMR, though some subdivision of these categories will at times be necessary. These sections will be interspersed with guiding questions in order to sharpen the focus of what is to follow.

2.1 Cambridge Language Testing and the FCE

At this early stage it is important to establish the basic constructs and concepts used in this research. With this in mind, the FCE will be outlined before cultural validity is explained.
A Brief History of the FCE

The FCE has its roots in the Lower Certificate in English (LCE), which was introduced in 1939. This exam was an intermediate-level qualification that acted as a precursor to Cambridge’s Certificate of Proficiency in English (CPE). True to the needs of the time, the LCE provided a means of regulating English language proficiency in an increasingly diverse wartime workforce (Hawkey and Milanovic: 2013). More importantly perhaps, the necessity of a common language meant more than just the establishment of a single linguistic code. It became a means of binding together disparate nationalities under a shared purpose, ethos and culture (Anderson: 1983, Kymlicka: 1995). At this stage the exam consisted of an oral paper, a composition paper and a choice of either a 'prescribed' British literature or a direct translation paper. The latter soon expanded to include eighteen languages to cater to the many backgrounds of its candidates.

This format remained in place for well over two decades until the 1960s and the early stages of the Communicative Approach. Notably influenced by the work of Arthur King on the UCLES / British Council joint committee, efforts were made to divorce English language learning from the literary and cultural knowledge that had characterised early versions of the exam (Weir et al: 2013). The avoidance of 'cultural propaganda' (Smith: 2004) and a growing appreciation of Britain’s increasingly marginal role in world affairs were central to this pursuit of a fairer 'purely linguistic-type examination' (Joint Committee: 1962 p.17). Consequently, a more communicative focus on 'general meaning' was adopted. Direct translation was dropped and 'General English' papers (lexis and grammar) were preferred alongside reading and listening comprehensions. Composition remained in place but the speaking paper took on a more reciprocal 'interview' format.

The name 'First Certificate in English' was adopted in 1975 and further revisions were introduced in 1984 and 1996 in which specific 'micro-skills' such as understanding gist/detail and identifying attitude were brought to the fore (Hawkey: 2009). With the growing prominence of the CEFR (Common European Framework of Reference) the FCE became identified with the B2 grade band (Council of Europe: 2001). In 2009 a version of the examination entitled 'FCE for Schools' was administered in tandem with the main FCE, designed specifically for teenagers and young adults. Finally, alterations were made to the format of the exam in January 2015, leading to the amalgamation of the reading and 'Use of English' papers and a resultant shortening of the overall exam by 30 minutes (FCE Handbook for Teachers: 2015). Current figures estimate that approximately 270,000 people in over 100 countries take the FCE annually.
2.2 Culture and Cultural Validity

Cultural validity is an essential consideration for any means of assessment that seeks to be both fair and appropriate for the world stage. Naturally, most of the FCE’s 270,000 candidates take the test in non-English-speaking environments and are proficient in a range of native tongues. Crucially, they are rooted in cultures that may vary wildly from one another in their affiliation with the British context. This represents a challenge to candidates who have little or no frame of reference for Anglo-centric cultural schemata and who could be marginalised by scenarios and representations that are fundamentally alien to them. This potential for bias is particularly acute during descriptive exercises as any candidate that has a cultural familiarity with the context depicted can necessarily provide greater detail if they have the linguistic skills to do so. For such candidates, the challenge is largely linguistic. However, for candidates that have a significant cultural barrier to overcome, the FCE can be seen as much more than a language test - but one that demands a deceptive level of integrated cultural knowledge. The extent to which the FCE promotes such representations will be discussed in the early analysis of this paper.

So, what exactly is meant by the term? At this stage, it is necessary to first establish a definition of ‘culture’ itself before focusing on how it relates to validity. ‘Culture’ is a notoriously amorphous and diverse concept to which even early theorists were able to attach in excess of fifty individual definitions (Kroeber and Kluckhohn: 1952). Depending on context, it has both objective and subjective meanings and can be found in the lexicons of disparate fields ranging from anthropology to aesthetics. For the relatively limited purposes of this paper, a stripped-down articulation of ‘culture’ is preferred here. Richards and Schmidt’s definition, suitable for use in an Applied Linguistics context, reads as follows; ‘the set of practices, codes and values that mark a particular nation or group’ (Richards and Schmidt: 2010 p.151). Clearly there is a degree of generalisation at work, though this definition is not employed here to oversimplify and essentialise the characteristics of its informants or participants. It is evident that, in a globalised world, many ‘cultures’ (and, indeed, subcultures) can be embodied by an individual at a given time. This is simply a working definition designed to capture the marked differences in ‘practices, codes and values’ between the UK (or, more generally, ‘the West’) and other world-regions, specifically the Middle East in this instance.

‘Cultural Validity’ refers to the extent to which a given test is ‘fair to the language and culture of individual test-takers’ and is characterised by the ‘avoidance of cultural bias’ (Cohen, Manion and Morrison: 2013 p.483). All tests are, by necessity, cultural artefacts and as such the production of a ‘culture-free’ test is an impossibility (Solano et al: 2011). Culture is an intrinsic aspect of a test, just as it is of its candidates. It cannot be removed in its entirety. We can however, seek to mitigate appeals to specifically cultural knowledge in tests designed for a global audience.

When referred to in this paper, the broader definition of bias should be understood as any item that ‘contains content or language that is differentially familiar to subgroups of test takers’ (Galaczi and ffrench: 2011 p.61, emphasis added)
2.2.1 Whose Culture? Whose English?

An anticipated objection to the above is that language and culture are inseparable and that, as a result, the latter is a valid area of assessment in a language testing context. Cambridge English may therefore, be justified in balancing concerns relating to cultural validity with a perceived need to assess what can be termed 'intercultural communicative competence' (Byram: 1997), though this is not an explicit goal of their examinations. Indeed, Cambridge allude to a degree of cultural specificity in their overview of the FCE, stating that the exam 'shows that you are ready to work or study in an English-speaking environment' (Cambridge English: 2011).

An important counterpoint to any contention that language and culture are purely symbiotic is that this attitude reflects a dated pre-globalisation world-view. It is now widely acknowledged that English is currently the most dominant international language worldwide (referred to as 'World English' or English as a lingua franca) (Crystal: 2003, Kirkpatrick: 2010, Jenkins et al: 2011). However, this growth in usage and influence is more profound than simply the recognition that English is widely used internationally. It represents a deeper challenge to the language's sense of 'location within' and 'belonging to' a set cultural context. Furthermore, it undermines English's canonical approach to lexis and grammar by acknowledging the diversity of its indigenised variants. These are known as 'World Englishes', as opposed to 'World English' (Crystal: 2003).

Though some earlier attempts to codify this spread of English do exist (Strevens: 1980), it was Braj Kachru's 'Three Circles of English Model' (Kachru: 1985) that provided the spark for what was to become an increasingly important field of research and commentary. The influence and cohesiveness of the English language was conceived of through three expanding circles. It can be seen to have a cultural-linguistic core ('Inner Circle') that dissipated in its strength of influence as it reached the circle's periphery ('Expanding Circle'):

![Image](image.png)

Adapted and updated from: 'Standards, Codification and Sociolinguistic Realism (Kachru:1985)

It is increasingly apparent that such a model does not accurately explain the sheer diversity of English is a post-globalisation era (Widdowson: 1997, Crystal: 2003). From the advent of the internet to the increased affordability of international travel, it has become extremely difficult to chart the exponential growth of
English and its various indigenisations. What has become clear however, is that the *reasons* learners engage with the English language have become equally diverse. As Jenkins puts it, ‘English spreads and adapts according to the linguistic and cultural preferences of its users’ (Jenkins: 2015 p.52). The goal of English language proficiency and the need to assimilate into an 'English speaking' country no longer map onto one another in the way they used to. On the contrary, considerable resistance to western cultural influence is evident in some world regions, particularly those with a post-colonial legacy (Bolitho: 2003).

Acknowledgement of the challenge this shift poses international assessment can be easily found in Cambridge's own research:

> 'English-language test takers in our contemporary context may have no common need to engage with, visit or work in the culture from which their English-language test originated.'
> 

Taken together, these observations have the potential to undermine international examinations that contain overt cultural bias. Whether these charges apply to the FCE will be discussed in the forthcoming analysis. In order for such analysis to be fair and grounded, it is important to address, on their own terms, the claims Cambridge attaches to the examination itself. This will be done with a particular focus on test fairness and global relevance.

**What are Cambridge English's Claims & Recommendations regarding 'Test Fairness'?**

**2.3 Cambridge's Commitment to Fairness and the Broadening of Cultural Context Initiative:**

Cambridge English attaches a number of claims to its First Certificate in English relating to its purpose, design, suitability and the evidence from which these derive.

**2.3.1 Overview of FCE Claims**

*Cambridge English: First shows you can use everyday written and spoken English for work and study purposes. The exam uses real-life situations that are especially designed to help you:*

- communicate more effectively
- learn the language skills you need

*The First Certificate in English:*

- provides the most reliable reflection of your language skills
- covers all major varieties of English
- is designed to be fair to users of all nationalities and linguistic backgrounds
- is supported by the largest research programme of its kind
- can be taken by people with a wide range of special requirements

Cambridge English: 2015a [bold type in original]

Cambridge English claim that the FCE represents a reliable and fair means of assessing English language proficiency at intermediate level and that these key contentions are grounded in rigorous academic research. Furthermore, it purports to be as fair as possible to candidates of all nationalities and linguistic backgrounds.
This latter claim is repeated in a number of guises throughout Cambridge's associated literature:

'[We]... ensure that our exams remain fair and relevant to candidates of all backgrounds and to a wide range of stakeholders.' (Cambridge: 2015b p.2)

'[Our exams]... are as fair as possible to all candidates, whatever their national, ethnic and linguistic background, gender or disability.' (ibid, Cambridge: 2014 p.2)

'Cambridge English exams are designed to be fair to all test takers.' (Cambridge: 2015a)

The straightforwardness and simplicity of these statements is perhaps understandable given their context in the Cambridge Handbook, a necessarily accessible publication. However, the level of academic rigour that underpins them is substantial. Cambridge English's background commentary and research in the field of language testing is, as they rightly state, unparalleled in its volume and scope. Clearly each individual paper of such a vast and ongoing body of work cannot always directly inform Cambridge's examination policy. However, there are projects conducted under Cambridge's auspices that, while not widely cited, are of critical importance to the cultural validity of their examinations.

One such project is the 'Broadening Cultural Context Initiative', which was tasked with producing a set of guidelines designed to ensure that examination materials do not 'contain a cultural focus that is too narrow, or [...] favour the views or assumed knowledge of one culture over another' (Murray: 2007 p.19). It makes the following key recommendations:

- It should not be assumed that candidates are knowledgeable about, or interested in, British culture.
- Texts and other task input material should not assume that candidates enjoy the lifestyles of particular income groups or nationalities.
- Exam materials should be drawn from as wide a range of sources as possible, including examples of non-British English.
- Good exam materials are often drawn from sources that have the widest possible original target audience.
- Where cultural assumptions might impede understanding, materials should be edited to gloss or, if necessary, remove cultural allusions or references.
- Assumptions should not be made about a candidate's location, cultural background and/or lifestyle.

'Broadening the Cultural Context of Examination Materials' (Murray: 2007 p.21)

With one peripheral exception (Wilson and Poulter: 2015), these guidelines have not since been endorsed by any subsequent Cambridge publication, nor for that matter, have they been repudiated. More specifically,
in the eight years since publication, no references to it can be found in Cambridge English's associated handbooks or guidelines for examiners. Similar recommendations had been made prior to the Initiative (Fried-Booth: 2004), but were reserved for advanced-level papers only. Consider the following example:

'...it can no longer be assumed that students learning the language know anything about, or have any interest in, British culture; if [...] a common cultural heritage [is assumed], certain cultures are bound to be alienated by these assumptions.' (IELTS/CPE Research Notes Fried-Booth: 2004 p.12)

Any cultural concerns relating to papers of an intermediate level or lower are subsumed under Cambridge's broader commitment to 'fairness' and, as such, are not well-defined. It is typical for the topic of culture to be considered peripherally as part of the discourse on test validity and not often dealt with head-on (Solano-Flores et al: 2011). The terms 'nationality' and 'ethnicity', despite their differing meanings, are often preferred as bywords.

Based on the above, Cambridge English's long-running and conflicted relationship with cultural content (Joint Committee: 1962, Smith: 2004, Weir et al: 2013) is not strictly one that can be consigned to the past given the nature of the recommendations made in areas of its own research (Fried-Booth: 2004, Murray: 2007). The extent to which these recommendations have permeated Cambridge English's FCE speaking papers will be returned to later in this paper.

2.4 Proposed Solutions:
Having outlined the FCE, the importance of cultural validity and the intersections between them, we now turn to two concepts that may offer a promising new perspective on test design and content. Essentially, these are proposed as means of reducing the need for heavily 'cultured' representations in a speaking test setting. The first, 'Learner Authentication', will form the philosophical basis underpinning this research. It will act as a guiding principle for the use of the second concept, 'Multidimensional Mental Representation', which will provide the practical means of 'operationalising' a more candidate-centred approach to the test. It is not the intention of this paper to present these as the only means of improving cultural content, for many such paths no doubt exist. They have been chosen primarily for their potential to centre testing material around the candidates themselves, by appealing to their personal lives and past experiences through the senses. If a test can be personalised in this manner, it could help circumvent some aspects of cultural bias within current test content, should they be found in evidence. Of these two concepts, it is 'Learner Authentication' to which we now turn.

2.4.1 Learner Authentication

What is Learner Authentication?
In conventional vocabulary, the word 'authentic' has an undoubted simplicity and resonance to it. It carries with it positive connotations, particularly when set against its antonyms of 'artificial', 'contrived' and 'synthetic'. Consequently, it is unsurprising that such a relatively crude bifurcation should dominate early efforts to engage with it in an ELT context (Sweet: 1899). However, 'authenticity' is a deceptively complex and controversial concept (Mishan: 2005, Fenner and Newby: 2005) that can change meaning when applied to different contexts. In ELT the term has since come to take on a tripartite form, broadly embodying the areas of Text, Task and Content.

For much of the 1960s and 1970s, textual 'authenticity' was widely considered to be any 'stretch of real language, produced by a real speaker or writer for a real audience' (Morrow: 1977 p.13). The rise of Communicative Language Teaching shifted the orientation of textual 'authenticity' towards a 'primary intent to communicate meaning' (Swaffar: 1985 p.17). Going further, 'authentic' texts should possess an intrinsic, natural communicative quality throughout (Lee: 1995 p.324). Similar themes run through the task-based use of 'authenticity'. Ellis proposed that the benchmark for 'authenticity' in TBL should be any 'task performed under real operating conditions in meaning-focused language activities' (Ellis: 1990 p.195). Finally, content authenticity holds that materials should be considered authentic if they 'fulfil some social purpose in the language community in which [they were] produced' (Little et al:1989 p.25). They should, in effect, be faithful to their original cultural context (Kramsch: 2000). Given that teaching and testing content are often culturally western-centric, and that they are designed for a range of audiences and purposes, this aspect of 'authenticity' has challenging implications for the field.

However, beyond this standard delineation between Text, Task and Content, we need more than simply a definition of the term. We also need clarity as to what it can or should apply to. Beyond the differing aspects of Text, Task and Content authenticity there lies a more profound distinction between pragmatic (product) and personal (process) authenticity. The former is often described simply as 'normal language behaviour in pursuit of an outcome' (Widdowson: 1990 p. 46). It is, in effect, an authentic 'action' resulting in an authentic 'product'.

It is however, the introduction of personal or process 'authenticity' (learner authentication) that forces us to re-evaluate our understanding of 'authenticity' (Fenner and Newby: 2000). In order to clarify what is meant here, let us expand upon Mishan's example of an archetypally 'authentic' learning experience - a young child's rapidly-expanding knowledge of his/her mother tongue (Chambers and Mishan: 2010).

What, precisely, is being referred to as 'authentic' here? It can be safely assumed that the child is 'authentic', not least because children lack the ability to consciously alter what they are and how they interact with their
environment. So then, our attention must turn elsewhere, to the myriad of linguistic stimuli that feed into his/her acquisition of language. Such real-world inputs are indeed, necessarily, 'authentic' in that they are neither artificial nor contrived. However, our analysis thus far lacks depth and is in need of expansion. What is truly 'authentic' here is the child's relationship with its learning environment (Chambers and Mishan: 2010). 'Authenticity' in this context is not a simple, static term. It is not a property of things, but one that exists between things (Breen: 1985, Taylor: 1994). It does not, and cannot, solely reside 'in' texts, tasks and test content but is reified through interaction and familiarity.

It is a key contention of this paper that international test content must be made accessible to the point of allowing for a workable level of learner authentication. This is simply because the relationship between candidates and certain types of exclusive test content (specifically the cultural in this instance) may vary greatly from person to person. This relational variation does not necessarily reflect linguistic ability, merely the backgrounds of the candidates themselves. As such, it poses a challenge to test validity and reliability, one that requires addressing wherever possible. Whereas this may seem a difficult task to achieve on such a large scale, it would become a distinct possibility if test content was orientated around the learner themselves. That is not to say that it should reflect the many cultures and aspirations of all who take it, for that would be an impossibility. Rather, it means that greater opportunities for personal reflection and self-expression should be opened up within the test itself. With a view to bringing this theory into practice, we now turn to 'Multidimensional Mental Representation'.

2.4.2 Multi-Dimensional Mental Representation

What do we mean by 'Multidimensional Mental Representation'?

Naturally, language is integral to who we are as individuals. It is the primary medium through which we govern our social and professional interactions. Our use of it defines what we are able to achieve, support and object to. It is integral to any comprehensive sense of volition and agency. More profoundly perhaps, it seems fundamental to what it is to be 'human'. However, whereas it is easy to assume that language is in some way 'innate', this does not map onto the best psychological, neurological and anthropological evidence (Damasio and Damasio: 1992). Higher-level primates and pre-verbal children can demonstrably grasp basic concepts triggered by external stimuli. So too, can diseases of the brain render adults able to identify simple constructs (such as colours) and yet be unable to correctly verbalise them (Decock and Horsten: 2000). It seems clear that there are basic cognitive preconditions for language that can exist independently of it. If language is a means of articulating sensory input and our mental representation of it, then such perceptions are primary and essential to its formation.

Our use of mental representation does not fade when we learn to use language proficiently. On the contrary, it continues to be the primary lens through which we interpret the world around us. It retains
its place as a vital intermediary between our sensory experience and the language we use to express them. As human beings we have been able to construct verbal representations of the world so 'natural', and yet so complex, that they obscure the layer of mental representations that underpin them (Damasio and Damasio: 1992, Masuhara: 1998).

Crucially, the relationship between stimuli and mental representation is a reflexive one. Linguistic and pictorial representations have the power to stimulate our sensory world even when the origins of such first-hand sensory experience are not present, but rooted in the past. It is for precisely this reason that we are able to experience visual, auditory, tactile and olfactory sensations when we view a picture or read an engaging text (Masuhara: 1998, 2005). These sensations are often accompanied by deeper affective engagement connected with the experiences and attitudes of the viewer/reader (Arnold: 1999, Tomlinson: 2000, Masuhara: 2005).

Simply put, MDMR is the way in which we internally process and visualise the world around us. The resultant representations are mental and therefore far deeper and more expansive than the 'raw data' received through the physical act of seeing. They are a unique, dynamic synthesis of external stimuli and personal experience. The reality of the outside world is filtered through the lens of our perception and we can 're-experience' this reality in a visceral way, provided appropriate stimuli are in place.

MDMR is a key component of what has since been termed 'multidimensional language learning' (Tomlinson: 2000), an approach that places the achievement of sensory and affective engagement centre-stage. Perhaps more importantly to the purposes of this paper, it also seeks to connect the learning process to 'previous experiences of language and life' (ibid: p.1). We see here a key element of personalisation at work, one that is hoped can be drawn on to help ameliorate cultural bias in test content by harnessing this process of personal 're-experiencing'.

Given its broad scope, MDMR (and, more generally the 'Multidimensional Approach') has a wide-ranging academic history and can be seen to overlap a number of disciplines. In order to gain a fuller understanding of this background, we will break it down into the following areas: experiential learning, neurological evidence and the use of multidimensional learning in ELT.
2.4.2.1 Experiential Learning

From a pedagogical perspective, MDMR falls under the auspices of 'experiential' education (Tomlinson and Masuhara: 2009). Pioneered through the influential work of John Dewey, experiential learning can be understood as a learner-centred approach that is driven by direct experience (Dewey: 1938) or 'learning through reflection on doing' (Patrick: 2011 p.1003). It views education as a personal, organic process rooted in a direct, interactive relationship with the focus of study. This process of 'learning by doing' is, in essence, one of subjective 'meaning-making'. Learners are involved in an ongoing experiential cycle that allows them to formulate and test hypotheses about the world around them.

Through time, this conception of the learning process was to influence a wide range of scholarly pursuits, spanning disciplines as diverse as philosophy, the material sciences, sociology, politics and cultural studies. Jurgen Habermas and Paulo Freire in the 1970s was particularly prominent in championing the transformatory potential of experiential learning (Freire: 1970, Habermas: 1972). Here the primary purpose of education was political - to elevate the critical consciousness of students by encouraging them to engage with the realities (and specifically the inequalities) of the world around them.

Despite its significant impact, it was not until the 1980s that rigorous efforts were made to codify the process of experiential learning into a practical pedagogical model. David Kolb's work in this area was key in this regard (Kolb: 1984). He proposed that successful experiential learning could be seen as progressing through four distinct stages. Concrete experiences (Stage 1) lead to reflective observations (Stage 2) which in turn give way to a period of abstract conceptualisation (Stage 3). Active experimentation (Stage 4) concludes the cycle, though multiple repetitions are necessary for durable understanding to take root. Experimentation provides new experiential data that fuels subsequent learning cycles. For Kolb, experiential input is educationally empty without learners reflecting upon it, just as abstract conceptualisation is ungrounded when divorced from its experiential base (ibid). Kolb's assertion that learners, as unique individuals, have strong innate preferences for specific stages of the cycle was also highly influential. These personal preferences denote their 'learning style', though the academic evidence for categorising learners along these lines is somewhat unclear (Dorneyi: 2005, Ellis: 2008).

It is almost impossible to imagine the world of education as it is today without experiential learning. It is the basis for everything from professional development to graduate internships to school trips. Inside the classroom it can be seen in the use of role-plays, discovery tasks, problem-solving, games, drama and a range
of other simulations. Many modern institutions often see past experience and education as interchangeable, offering to waive academic credits in place of real-world experience. The fact that this very dissertation seeks an evidential base in the form of a case study is itself informed by the importance of experiential learning. The influence such learning has had on education cannot be underestimated.

2.4.2.2 Neurological and Psychological Background

Mental representation has been the subject of much theorising and research in the cognitive sciences. From the first days of experimental psychology, mental representation was seen as integral to understanding how the mind works. As early as the late nineteenth century, William James wrote of visual and verbal imagery as being the base 'commodity' of thought (James: 1890). In the same vein, Wilhelm Wundt was among the first to claim that the very notion of what we conceive of as an 'idea' is in fact rooted in 'memory images' (Wundt: 1912). Crucially, these images comprised of the combination of inputs present during the formation of a past experience, not merely the first-hand perception and recognition of a given object or concept (Carpenter: 2005). In this view it is our perception of, and reflection on, the past that is integral to the very fundamentals of cognition.

Throughout the decades that followed such conceptions of mental imagery were seen as overly introspective and discarded in favour of more empirical, 'behaviouristic' approaches (Watson: 1930, Boring: 1950). In the 1950s however, there was a marked shift away from the behaviourism of studying strictly observable responses to stimuli towards investigations into how the mind combines and represents the inputs it receives. It became increasingly apparent that we often encode the information we receive into indistinct units (Miller: 1963), mental representations that carry with them a wide range of sensory data. Once sets of stimuli are synthesised into these 'chunks' of information over time, they become established as powerful and durable mental constructs. Such mental representations appeared not to be subject to the same memory constraints that act upon the individual inputs that comprised them (ibid: 1963). Furthermore, words which directly connect to the corporeal world (and therefore can be easily mentally visualised) proved significantly more likely to be memorised than abstract terms (Paivio: 1965). This formed the basis of Paivio’s 'dual-coding' theory in which both visual imagery and verbal associations were key for acquiring and consolidating new information. It was claimed that when the two inputs combine the resultant representations are more accessible and easy to 're-experience' when triggered (ibid).

In the 1970s attention turned to the properties of mental representations. Kosslyn advocated a quasi-pictorial view, claiming that they share the spatial and graphic qualities of the observation(s) from which they are drawn (Kosslyn: 1973). The mind constructs an archetypal 'surface representation' as a buffer and
then subjects the resultant image to analysis and interpretation. His contemporary, Pylyshyn, preferred a more propositional approach in which 'spatial' images are incidental to the core processes that accompany mental representation and that symbolic or abstract imagery can theoretically fulfil the same role (Pylyshyn: 1973). In this view, the propositions behind a particular image take cognitive precedence over the images that accompany them (ibid, Anderson and Bower: 1973).

Moving into more contemporary research, advances in technology opened up new avenues through which to explore the physical processes that accompany mental representation in brain. True to Paivio's hypothesis, the evidence strongly indicates that verbal and visual inputs share psychometric properties but that they demonstrably stimulate different parts of the brain (Mazoyer et al: 2002). However, the resultant mental representations (whether created by either visual or verbal stimuli) have the same psychophysical properties when recalled (Denis, Goncalves and Memmi: 1995). Once the information is received assimilated and subsequently re-experienced, it appears to be treated in broadly the same way. This leads us to believe that, regardless of the stimulus in question, mental representation is fundamental to even the simplest acts of cognition (Thompson and Kosslyn: 2000). Perhaps most interesting of all is the evidence that suggests the brain is more likely to construct durable mental imagery if the input is solely visual or solely linguistic (Mazoyer et al: 2002). This contradicts much of the early theorising in this area and supports pedagogical claims that detailed visual accompaniments to text can disrupt the creation of mental representation (Tomlinson: 2013).

Whereas the psychophysical underpinnings of mental representation remains a relatively new and contested area of study, there is certainly sufficient evidence at present to support the integral role such representations play in the learning process. These findings point towards experience and recollection of experience as both possessing similar properties and being capable of achieving similar results. More specific to the purposes of this paper, they show that such experiences can, in part at least, be recreated while maintaining their capacity to animate the senses. If highly-engaging and affective past experiences are to be used a basis for personalised test content then this link is very important as it indicates that such 're-experiencing' may be possible in a closed setting, if conditions are conducive to it.

2.4.2.3 MDMR and the 'Multi-Dimensional Approach'

Perhaps due its technical nature, Multi-Dimensional Mental Representation does not currently have a large body of literature associated with it in the ELT field. Direct references to it do not date back further than twenty years (Tomlinson: 2000, Masuahara: 2005), with Masuhara's 1998 unpublished thesis being the first
known expression of the term in this guise (Masuhara: 1998). However, MDMR is alluded to in ELT through a number of associated terms and theories. All mentions of 'multisensory learning' and even simply 'mental imagery' (Richards: 2010) at least in part motion towards an appreciation of MDMR's importance.

The term can be seen to draw together a wide range of disparate realisations, some of which have been present in the field for decades. True to the progressive nature of the era, the 1960s and 70s proved a vibrant time for experimenting in this area of the field. However, many novel approaches to learning focussed heavily on achieving SLA through a limited conception of sensory perception via appeals to specific senses. We can see the auditory focus of the Silent Way (Gattegno: 1963) and Total Physical Response's strong preference for kinaesthetic learning (Asher: 1965) as being indicative of this restricted appreciation of sensory potential. A more expansive (though still relatively limited) appeal to the senses can be seen in the form of Suggestopedia, which incorporated eyes-closed reflection on language input, rhythmic breathing and periods of both silence and classical music (Lozanov: 1978, Bancroft: 1978). Whereas Suggestopedia did not attain mainstream prominence in the classroom, there were related theories from this period that went on to achieve more influence in ELT and education in general. Notable among these were the Humanistic Approach and Neuro-Linguistic Programming.

The Humanistic Approach viewed education primarily through the lens of personal growth. Strongly advocated for in the work of Gertrude Moskowitz, it recognised the importance of the affective dimension of learning, as well as the cognitive (Moskowitz: 1978). Consequently, the subject matter taught was heavily learner-centric and combined with the feelings, experiences and the lives of students (ibid). This increasing importance attached to affective engagement was further evidenced in the influential work of Steven Krashen, in which the lowering of the 'affective filter' was given a prominent role as one of his five key hypotheses of second language acquisition (SLA). The link between affect and durable SLA has since been widely corroborated, both from psychological (Damasio:1994, Schumann:1997, Immordino-Yang and Damasio:2007) and pedagogical perspectives (Arnold: 1999, Tomlinson: 2000, 2013, Tomlinson and Masuhara: 2009).

In contrast to humanistic approaches, Neuro-Linguistic Programming (NLP), sought to recreate successful patterns of behaviour by 'recoding' the way in which learners subjectively respond to stimuli (Bandler and Grinder: 1976). Experiences are encoded through the senses (vision, tactition, audition, gustation and olfaction) and have a discernible structure that can be replicated and re-experienced in pursuit of optimal behaviour (ibid). Despite some notable advocates (Revell and Norman: 1997, O'Connor and McDermott: 1996, Rinvoluci and Baker: 2005), NLP is heavily criticised from many quarters for being anecdotal and pseudo-scientific (Langford: 1999, Thornbury: 2006, Bovbjerg: 2011). However, as a personal philosophy or justification for already-established teaching techniques NLP has been considered by some to possess a degree of merit (Richards and Roberts: 2001) [though this assertion has been removed in the 2014 edition of this title].

Another, perhaps less controversial, appeal to the importance of multidimensional learning can be found in
the form of 'holistic' approaches to education. The term 'holism' is used by many in ELT to refer to a 'focus on language in its entirety rather than.. [its] separate components' (Richards and Schmidt: 2010 p. 262). Language should be taught in its natural, integrated form and loses many of its key communicative features when atomised into its discrete parts. However, 'holistic approaches' are also used in reference to 'whole person' learning; the 'holistic and multisensory nature of learning that involves head, heart and hands' (Legutke and Thomas: 1991 p.159, emphasis added). Language learners, as multifaceted individuals, must be taught in a manner that appeals to and reflects their complex nature. Consequently, the 'cognitive, affective, emotional and social' needs of students are all important considerations in effective second language acquisition (Goh and Burns: 2012 p.4).

The purpose of the above chronology is not to lend credence to all theories extolling the educational benefits of sensory perception and emotional response. Rather, it represents an attempt to chart the field's ongoing fascination with such input and the harnessing of its learning potential. Looking over the contributions, we can see a growing realisation of the complexity of the learning experience and, indeed, the complexity of perception itself.

So, where does the multidimensional approach stand in relation to what has preceded it? The multidimensional approach, of which MDMR is a key component, can be seen as situated within this same appeal to the 'whole person' or 'whole learner'. As outlined by Tomlinson, it facilitates sensory and affective engagement while connecting learning to previous personal experiences of language and life itself (Arnold: 1999, Tomlinson: 2000). It seeks to ground its claims in interdisciplinary research in a more rigorous way than many of its theoretical forebears (Masuhara: 1998, 2000) and typically employs a range of mental imaging, inner voice, kinaesthetic, language awareness, and cognitive processing activities to achieve its goals (Masuhara: 2000, Tomlinson: 2000). We must acknowledge that it may not be possible or appropriate to use all such activities in an examination setting while remaining faithful to the FCE's core purposes. However, it is hoped that particularly imaging and reflective activities will assist in creating a testing environment conducive to the recall of strong mental representations that will aid self-expression.

Chapter 3: Methodology

3.1 Overview and Rationale of this Research

It is the intention in this paper to explore culturally-sensitive alternatives to the FCE speaking paper by trialling items designed to be more personal and accessible to as wide an audience as possible.
Drawing on the literature review above (most prominently the work of Mishan, Tomlinson and Masuhara), these alternatives will be drafted with Learner Authentication and opportunities for Multidimensional Mental Representation in mind. These have been chosen for their potential to create spaces within the examination that allow for the generation of content that the test-taker necessarily has a prior familiarity and relationship with. Essentially, the articulation of past mental representations will form the basis of candidate output, thereby sidestepping the need for pictorial representations that may be exclusive and culturally engendered. The fact that such an approach could result in fairer examination content and improved performances forms the driving hypothesis behind this paper.

This dissertation is seen as primarily qualitative. Whereas it does appeal to Cambridge's objective rubrics of assessment (Appendix H), it is more concerned with, and directly informed by, the perceptions and feedback of its participants. It is exploratory in nature and seeks to use qualitative methods towards nominally quantitative ends. These ends are the results gleaned from the original and alternative test items, whereas the qualitative background input is obtained via interview and general comment. A loose interview format, as opposed to a more rigid questionnaire, was chosen for its directness of contact with the participants and its capacity to add nuance and personal input to a potentially complex area of cultural commentary. The drawbacks of this approach will be acknowledged later in this chapter.

The first part of the trial will consist of FCE test-items adapted from the original tests and administered under as close to examination conditions as possible. These results will act as a baseline or control group against which to judge the changes in candidate response. They will be assessed against Cambridge English's marking rubrics and summed up with commentary on the candidate's grammar, vocabulary, output, pronunciation and fluency. In line with this paper's own focus on personalisation, analysis of the items' ability to connect with the candidates will also be subjected to scrutiny, in the form a 'content engagement' section. The second part will focus on the alternative test-items. The judgements that relate to the appropriateness of the original test content will be drawn primarily from the feedback of the candidates themselves. The replacement items will be informed by the above research, with a particular focus on the aspects of the Multidimensional Approach that aid MDMR (eyes-closed activities, silent reflection and affective engagement) (Tomlinson: 2000). These were chosen from a wider pool inspired by MDMR-maximising classroom activities (ibid, Tomlinson: 2013) and the selections were based on the ease with which they could be implemented both in a constrained 'examination' setting and within a research project limited in scope and time-frame. For comparative purposes, candidate performance in these items will be assessed and analysed in the same way as above.

3.2 Structure of the FCE Speaking Paper

This paper is concerned only with those portions of the examination that require the candidates to engage with 'in test' visual imagery. Whereas it would be possible to analyse bias in the textual aspects of the FCE (particularly the writing and reading papers), such discourse-driven analysis lies beyond the scope and
time-frame of the current research. With this focus on imagery in mind, we now look at the structure of the speaking paper in order to highlight the areas that best meet these criteria. The FCE is currently spread over approximately 15-20 minutes, comprising of four equally-weighted parts:

1) **Personal Questions** – Provides opportunities to offer personal information and appropriately respond to questioning. The examiner asks the candidates individually simple questions about their own lives. These questions are themed around daily routine, work, study and interests.

2) **Long Turn** – Provides opportunities to impart information and express opinion. The examiner gives the candidates two pictures each and directs them to their accompanying question. They answer their own questions and are permitted to comment on their partner’s afterwards.

3) **Conversation** – Provides opportunities to impart and exchange information, attitudes and opinions through collaboration and negotiation. Candidates are given a set of related photographs and asked to discuss them.

4) **Discussion** – Tests candidate ability to engage in a discussion, based on the topics or issues raised in the collaborative task. The candidates are asked to respond to questions, initially relating to their conversation about the photographs. This discussion is gradually widened in scope by the examiner.

The 'Personal Questions' stage was trialled and excluded due to an absence of visual imagery and the fact that the questions were deemed sufficiently personal in nature. Candidates did not struggle with the item’s content (e.g. 'What do you usually do in the evenings?', 'What do you like doing on holiday?') and their post-test feedback indicated little or no confusion on cultural grounds. Output was measured at 115 words per minute, with a broadly even split between the two candidates. A full transcript of this trial can be found in the appendix (Appendix B p.63).

The following three items are, to varying extents, image-driven and containing between five and ten pictures between them, depending on the year. Consequently, pictures relating to these stages will be the focus of this analysis.

### 3.3 Participants

An Arabic-speaking intermediary assisted in identifying and approaching suitable participants studying their foundation year. This was done in public through their institution’s central library. The criteria for selection were as follows:

1) Participants must be, or at least self-define as, culturally Middle-Eastern.
2) While test participants must be UK-based, they would have limited prior exposure to western culture and ideally would not have been in the UK for long enough for acculturation to have taken place.

3) Given the constraints some Middle-Eastern countries place upon gender-mixing, participants should be male in order to minimise any possible reticence or discomfort attributable to gender differences.

4) Participants should be of a broadly intermediate level of English language proficiency.

5) Participants would be in an age-bracket that made them eligible for both FCE and FCE for Schools.

6) Having been fully informed of the research, all participants must freely give their ethical consent.

The research was eventually built around the contributions of two young Kuwaiti nationals, both male and aged twenty. They shall henceforth be referred to as Candidates A and B. As candidates with an IELTS mark of 6.5, their proficiency had demonstrably been graded at B2 level. However, initial impressions suggested that their current levels may be marginally lower. At the time of contact, each had spent less than a year in the UK. Both indicated that they didn't know the country well and found it 'very difficult to make English friends' (Appendix B p.63). Whereas neither a thorough vetting process nor a large dataset were possible due to time constraints, the participants did meet the selection criteria and were considered broadly representative of the intended target group. Candidates A and B kindly agreed to the interviews and tests that made this research possible. The process this clearance was obtained through will now be detailed.

3.4 Procedure

1) Obtaining Ethical Clearance

Given the learner-centred and affective nature of this dissertation, ethical considerations were paramount from its earliest stages. The research design and content underwent some significant changes. Initially, it had a broader scope that had the capacity to elicit a range of affective responses. This potentially allowed for responses from the negative end of the emotional spectrum and was subsequently abandoned on grounds of failing the non-malfeasance test. Some of these items are outlined in the 'Suggested Alternatives' section of the analysis. A positively-orientated approach was eventually preferred, one that openly instructed candidates to create and engage with subject matter that they felt most comfortable with.

An information sheet (Appendix A p.62) was provided to both test candidates prior to them reading and signing the accompanying ethical consent forms (Appendix I p.72 [Omitted due to British Council rules on branding]). Both of these had been pre-approved by the University's ethics board. A fully bi-lingual (English-Arabic) speaker was made available during the explanation of ethical consent in order to answer any questions the participants had regarding their participation, right of withdrawal, anonymity or the complaints procedure. No such queries were forthcoming, though the bi-lingual intermediary was asked, by myself, to provide a brief overview of the consent process in Arabic. He remained present throughout the clearance process to help answer any subsequent questions, though this ultimately proved unnecessary.

A further important ethical consideration was that candidates and informants had free-reign to choose the timing of the test and subsequent feedback. Whereas this is desirable in any research situation, it was made
all the more acute by the fact the test candidates, and the intermediary, were all Muslim. As such, they were engaged in longer periods of fasting and prayer for the holy month of Ramadan. Sometimes late-night meetings were preferred and sometimes daylight hours suited their schedules, but the timing of these meetings had to be firmly geared towards the convenience of the participants themselves.

2) Obtaining Background Perspective
Despite having experience of the Middle Eastern learning environment, as a westerner I was only able to give anecdotal accounts of the FCE’s cultural deficiencies. It would be antithetical to the principles driving this research to suggest that my own views alone were sufficient in this regard. Consequently, the candidates themselves were interviewed in relation to the representations presented in a range of FCE papers. This stage extended to the trialling of FCE papers in order to obtain retrospective feedback. The tests provided for discussion and comment were all relatively recent papers approved by Cambridge English, dating 2008 to 2015. The purpose of this phase was to gain a more authentic understanding of the test from a differing cultural perspective.

3) Administration of the FCE Items
FCE test items were administered to the two candidates in a quiet room of their university's central library. The ambient noise was minimal, making it as close as possible to examination conditions. As is standard, the test divided into four parts, timed at an average of four minutes each. Allowing for examiner talk-time and 'picture viewing' time, this filled the requisite twenty minutes. Due to the conversational nature of the test, a leeway of twenty seconds was given to the recommended times per item (Cambridge’s rubrics allow for such measures to be taken). A range of test items were selected, dating from 2008 to 2015. Once the test had been compiled it was administered in line with Cambridge best practice.

4) Candidate Feedback
Candidates were asked to discuss and comment on the FCE, with a particular focus on its visuals. A wide range of FCE picture items (numbering approximately thirty, and later expanded to fifty) were placed face-up on the desk (Appendix G). Again, these dated from recent Cambridge-approved FCE exams and sample papers dating from 2008 to 2015. The discussion was designed to be free-flowing and loosely structured, guided by simple questions such as 'Where do you think this picture was taken?', 'Where are these people from?' and 'What do you think they are doing?' (see page 30 for further examples). These questions were designed to initiate cultural commentary because the merits of hiding their purpose were lost as both candidates had necessarily received an overview of the research during the ethical consent process. The results of this feedback can be found on pages 30 through 34.

5) Drafting of Alternative Test Items
Informed by candidate feedback and performance, as well as the background perspectives collected in this research’s earlier stages, a range of alternative test items were considered. The alternative items were designed by myself as the researcher and themed around the provision of improved opportunities for MDMR and learner authentication. They were loosely based on exercises and activities used in the
Multidimensional Approach (specifically eyes-closed activities and silent reflection) and sought to remove the perceived need for visual representation completely, or simplify its usage, in the hope of providing a more fair and neutral test format.

As previously mentioned, not all drafted test items reached the implementation stage. Ethical considerations (principally the potential for negative affect) played a major role in ruling many items out, though practical concerns (such as clarity of instruction and in-exam time constraints) also played their part. A full overview of the items considered for implementation can be found in the analysis contained in the following chapter (4.4 p.33).

6) Administration of Alternative Test Items
The test items that made it through the selection process were administered under the same environmental conditions as the FCE items (see point 3 above). Whereas the items themselves naturally differed in structure and content, their academic purpose and the time allowed for them were broadly the same. This was done to make the comparison between the alternative items and the original ones as fair as possible. In line with the principles guiding this research, a number of key alterations were made to the administration of the alternative items. Most notable among these were the addition of periods of silent reflection, eyes-closed responses and a more open role for the examiner.

7) Test Marking
Both the original and alternative FCE tests are marked against the same criteria. These criteria have been taken directly from Cambridge's own assessment rubrics (Appendix H). To maintain objectivity in the marking process, I took no part in it other than to analyse the results.

8) Analysis of Data and Discussion
Drawing on the research outlined previously, this stage documents the feedback gathered on test content and highlights any recurring themes therein. It details some viable alternatives to the more problematic sections of the test, focusing on those most appropriate for implementation. The candidate performance and feedback are then transcribed and subjected to scrutiny. Possible improvements on (and expansions of) this line of research will be considered during the conclusion (p.52). Recommendations for FCE test writers are provided as final comment (p.55).

3.5. Research Limitations and Concerns
As with all such projects conducted at this level, this research is limited in its scope and time-frame. It does not seek to provide statistically significant results but can hope to offer indicative evidence of MDMR's viability as a basis for international assessment, evidence that may warrant further investigation.

With regards to the examination guidelines laid out by Cambridge for the FCE, it is acknowledged that a
degree of examiner discretion is in-built so as to avoid the testing of inappropriate content. Examiners have a choice as to which visuals are used for certain sub-groups of candidates, a choice that potentially undermines the premise of this paper. The value of this element of cultural discretion is difficult to codify, given that the results of the choice(s) made are only evident after the fact. Additionally, the corpus of approved visuals this choice is made from may need to be uneasily large to sidestep such bias. This paper seeks to explore the potential for cultural bias in any and all such visuals. The prospect of testing candidates using inappropriate visuals shouldn’t be a mistake that such an examination allows, even as a possibility, particularly if alternatives can be offered.

Another early concern revolves around the specific role the FCE visuals play in an examination setting. They can be viewed merely as peripheral prompts that are subordinate to written rubrics which are generally less susceptible to cultural bias. This is not a position shared by this paper, chiefly because the visuals are thematically important to setting the scene for much of what follows their introduction. Even if their function was simply that of ‘prompt’ or ‘springboard’, their importance is in no way diminished on these grounds.

It is accepted that any such examination is required to present content that is, at least to an extent, unfamiliar. That is to say, candidates should be fairly challenged by the content and have no prior specific knowledge of the items included. Unfamiliarity also provides valuable opportunities for speculation, a key language skill for the B2 grade-band. However, to use the provision of opportunities to speculate as a means of circumventing the cultural concerns of this paper is questionable. Whereas candidates with a wider cultural scope can rely on this knowledge to aid their fulfilment of the examination’s more descriptive components, those without such a familiarity of context could be forced to forgo these easy marks and speculate straight from the outset. Stronger candidates will be able to do this, though weaker ones overrelying on speculation may struggle. Speculation is a higher-order language skill than description. It is unreasonable to put certain sub-groups of candidates in a position in which they are required to use it more than others.

The necessary ethical constraints have been addressed in the previous section and will be further detailed in the subsequent discussion. Further to these, it should be added that ideally candidates would have been tested in their home countries to ensure absolutely minimal exposure to western culture. Due to this approach’s unfeasibility on grounds of ethical consent and practicality, candidates were required to be UK-based. However, those selected did fulfil the personal criteria necessary for them to participate in a meaningful way.

Additionally, several rounds of testing and interviews would have been desirable to re-enforce the data collected. This would allow for the test to be refined over many iterations and for a larger, more diverse pool of candidates. Furthermore, increases in detail and volume of the collected data would make it easier to identify the processes at work and avoid false attributions of causality. As things stand, the connection between the MDMR items and any resultant improvements in performance would require a great deal more supporting evidence before firm conclusions could be drawn. However, there was insufficient time to conduct multiple repetitions of the tests in a manner that would not be excessively demanding on candidate time and energy.
The interview format used to collect bulk of the data is also potentially problematic on methodological grounds. Whereas it may seem the most logical way of obtaining direct feedback on a specific aspect of test content, interviews remain a contested means of eliciting such information. They can be leading and manipulative, to a point where participants cannot but voice the specific 'truths' that are congruent with the research's hypothesis (Silverman: 2001, Potter and Hepburn: 2005). In this view, all data obtained via interview is, in large part, 'manufactured' by the researcher. Furthermore, interviews arguably represent a much more familiar discourse setting to 'western' participants than to those rooted in other cultures (Litosseliti: 2010). It is for these reasons that the interviews in this research were loosely structured, open and conversational in nature. Candidates were encouraged to expand on their comments where necessary and no time limit was imposed on the discussion. Guiding questions tethered to the purposes of this paper were necessarily integral to the focus of the interviews, though these were not designed to manipulate or create false consensus. The purpose and context of the interviews was known to all, informed consent having been provided at the outset as a matter of sound ethical practice. The decision to inform had to take precedence over more peripheral methodological concerns. It is accepted that this may, at least in part, have unavoidably compromised the responses obtained. We have no guarantee that the known context of the research, combined with the complicity of the participants themselves, would not work together to produce overly-favourable results (participants giving what they believed were 'correct' or 'desired' responses). That having been said, there is no such thing as context-free research. Provided they are appropriately set up, interviews are arguably no less contaminated by their driving purpose than other forms of data gathering (Litosseliti: 2010).

Finally, this paper cannot satisfactorily take into consideration that the ability to imagine and reflect is likely to vary greatly from candidate to candidate. No guidance exists that can adequately account for or mitigate this possible disparity. Clearly more research would be necessary to ascertain whether such a focus on mental representation is, in itself, inherently unfair. However, such abilities are not fundamentally different from the range of other 'individual differences' that examiners generally have to ignore (Dornyei: 2005). The participants in this study will be judged against their own performance in the standard equivalent FCE test items, so in the context of this research they are in competition with themselves, not each other. So long as improvements are consistent between the tests such concerns are not considered detrimental to the value of this paper.

Chapter 4: Analysis

4.3 Background Perspectives

As outlined in the methodology, the first stage after obtaining ethical clearance was the gathering of candidates’ perspectives on the FCE’s visual content. A range of images were taken from the various recent FCE speaking papers dating between 2008 and 2015. These images included those used in the 'FCE for Schools' papers, due to their identical format and the fact that the candidates fell between the average ages of both examinations. For reasons of fair coverage, there was no strict selection process at work here other
than the dates of the papers. All photographs that could be found were considered viable for commentary, including those from Cambridge-approved supplementary materials. However, some preference was given to photographs from actual past papers and exemplary examination samples used in the Cambridge Handbooks. These photographs were initially compiled and organised into a portfolio of approximately thirty pictures, which was later expanded to fifty (see Appendix G). With the picture sheets evenly spread evenly across the desk, the participants were asked to spend time looking at them, provide general comments and answer some guiding questions.

**Picture Portfolio (Sample)**

4.3.1 **Candidate Comments**

“I can't say anything is so important, but ...that one I. I don’t have any idea about, what's that actually? [shrugs shoulders] So I can't say if it's important or not. [...] in my culture, in my countries – we never ever have it”

Candidate B on ‘Drama Picture’ in ‘School Activities’

As evidenced by this particularly fitting excerpt, feedback from the candidates proved extremely useful in highlighting problems in the general cultural orientation of the picture portfolio. Conversation was free-
flowing and lasted approximately 20 minutes. The candidate responses given during the test will also be called upon in this section as commentary where relevant.

The guiding questions, designed to elicit loose cultural comparison, were necessarily short and straightforward. They took the form of the following, and subsidiary questions were asked where appropriate:

<table>
<thead>
<tr>
<th>Question Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Where do you think these pictures were taken? Which country or countries? Why?.</td>
</tr>
<tr>
<td>What activities did you do at school? [in reference to the school pictures]</td>
</tr>
<tr>
<td>Who in these pictures do you think is richest? Who do you think has the least money? Why?</td>
</tr>
<tr>
<td>Are there any photos that you find confusing or don’t understand? Why?</td>
</tr>
</tbody>
</table>

4.3.1.1 Initial Comments

Below are selected excerpts from the opening of the discussion, in which candidates are asked for their initial impressions of the location and content of the pictures.

**Researcher:** I’m going to now show you all the photos from this Cambridge Exam. So, there are maybe, I think, just under thirty photos in total. Take a look at whichever ones you want.

[candidates take some time to look at the photo sheets spread across the desk]

*just want to know, where do you think these pictures were taken? Which country, or countries?*

**Candidate B:** [gesturing towards sheets 2, 3 & 4] *England, England and the UK. Absolutely in the UK* [now points specifically to ‘policeman’ picture on sheet 4 – Figure A]

**Researcher:** So that’s the policeman? Yah, I agree

**Candidate B:** *That’s the UK as well* [now points back to sheet 2 – Figure A]

**Candidate A:** [nodding] Yeah, some. This is somewhere in Africa I think [points to ‘tribal music’ picture above it on sheet 1 – Figure C].

**Researcher:** This? Yeah. What are they doing? I don’t know what they’re doing.

**Candidate B:** Dancing.

**Candidate A:** I think it’s typical festival or something like that in Africa.

**Researcher:** Yeah, okay... So, this is Africa, [points to Figure C] this is England [points to sheet 2], this is England [points to sheet 4]

**Candidate B:** *This is England* [points to sheet 4]... *this is England* [now points back to sheet 2]
Researcher: *What about these pictures here?* [points to sheet 5 and 6]

Candidate B: I can't tell you. They could be anywhere, but.. *when I see the pictures.. I imagine.. I think... that they are English.. British. I think* [points to 'Camping Family' – Figure D]

Researcher: The Family?

Candidate B: [nodding] *These as well* [points to 'Picnic – Figure E]. *They would never ever be in the Middle East or Arabic Countries. They look like 'English'*

---

As evidenced above, Candidate B felt strongly that the bulk of the pictures did not reflect a wide range of nationalities and cultural backgrounds, with Candidate A showing signs of agreement. With the exception of the African 'tribal music' photo, which itself conforms to an number of essentialist cultural stereotypes (Hall: 2003), the candidates' initial impressions of the portfolio were that it was almost exclusively Anglo-centric in the 'real-life situations' it purported to represent (Cambridge English: 2015).

### 4.3.1.2 Specific Comments

Both candidates were also asked to comment on some specific examples. Notable among these were the
'School Pictures' (all Sheet 3), 'Family Outside House' (Sheet 1) and 'Injured Footballer' (Sheet 4). We will now look at excerpts from these three parts of the conversation.

'School Pictures'

Researcher: Do these pictures look like your own experience of school? Did you do all of these things at school?

Candidate B: We do that [points to 'Computer-room Picture']. We do use computer rooms. We have, errr like.. three lessons, three or four lessons a week.

Candidate A: [points to 'Drama Picture', Figure F] Only this one.. because..

Researcher: What is this one? Yes..

Candidate A: In Arabic culture this, I think, is connected with 'Chinese culture'. It's kind of dancing or something like that.

Researcher: Mmmm..

Candidate A: [unintelligible] We don’t have this ‘thing’ in our school [points to 'Cooking Class', Figure G], but other things yeah.. and the ‘cook’ is not for male I think because..

Candidate B: It depends on the schools.. like, not all the schools they allow, y’know, to cook.

Candidate A: We have different things like music, we have ‘game' and drawing. Other stuffs, not like this errr, ‘activity'..

Researcher: Which two activities do you think are not so important?

Candidate A: Not important?. I think singing and dancing.

Candidate B: Actually all of them I think is important. I can’t say anything is so important, but that one L.. I don’t have any idea about, what’s that actually? [shrugs shoulders] So I can’t say if it’s important or not.

Researcher: I think maybe it’s acting or theatre.. or ‘drama'.

Candidate B: It might be, but in my culture, in my countries – we never ever have it.

As is evident above, some of the 'School Pictures' proved problematic and neither candidate was able to fully relate to their content. They particularly struggled with the idea of drama as a subject of study or extra-curricula activity and this would no doubt disadvantage them in an examination setting in which candidates
with a more western frame of reference would be able to further expand. So too, did the 'Cooking Class' picture (with the boy foregrounded) challenge Candidate A’s conception of gender roles (‘the ‘cook’ is not for male, I think’). It is not clear whether the candidate believes that the 'male' should not participate in this manner, or if he simply doesn’t recognise the scene as indicative of his own experience. It does however, show a disconnection between candidate and content that is more pronounced than it would otherwise be in a western setting.

The question 'Which two activities do you think are not so important?' could not be successfully answered by Candidate B. The reasons for this, as he acknowledges, were primarily cultural. He had no frame of reference for making an informed decision as to the activity’s worth and was unable to comment further. In so doing, he provides us with perhaps the most clear and open comment on the inhibiting effect such a disconnection can have on performance.

'Family Outside House'

Candidate A: [pointing at ‘Family Outside House’, Figure H] I think they are Americans because I think this kind of home is [unintelligible] like American house and the streets, we don’t have house like that.

Researcher: Okay, so this is not.. this is maybe also America?

Candidate A: Because they have hot weather like in Kuwait. This is not a very normal home but it’s errr, the home is from wood and it’s not typical, not typical. home like in Kuwait.

Aside from the disconnect between the picture and Candidate A’s own conception of ‘home’, this example is notable for three further reasons. Firstly, no effort was made to identify what the candidates are doing. This test item is themed around ‘family activities’, with boardgames the focus of this particular picture. However, the house was what the candidate was drawn to. Secondly, this was among only three pictures that the candidates singled out as indicative of poverty, the others being 'Tribal Music' (Figure C) and 'Biker' (Sheet 4). Interestingly, they contrasted this with the wealth they associated with 'Businessman Lunch' and 'Couple Driving' (Figures I & J, overleaf), which points towards a further possible bias in ethnic representation. Finally, much like many other non-white ethnic groups depicted in the other photographs, these people are identified by the candidates as ‘western’ (in this case American). A range of ethnicities are indeed present in the portfolio, though they are often set against what the candidates feel are specifically western cultural backdrops.

Figure I – Businessman Lunch   Figure I - ‘Couple Driving’
In this excerpt, gender roles again come to the fore and both candidates easily establish the picture's content as culturally 'English'. Whereas the narrative behind the pictures is not lost on them, the 'real-life situation' portrayed is identified as a quintessentially English one that does not map onto their own experiences. It is not clear whether this causes them any anguish or discomfort, though their ability to relate to the picture's content is at least partially compromised or distracted by its cultural differences.

4.4 Suggested Alternatives

Having established a number of cultural fault-lines within the FCE's visual representation, focus now shifts to the designing of more neutral test items that allow for a greater degree of learner authentication and MDMR.

During the early stages of this research, a range of alternative approaches were considered for piloting. Naturally however, not all made it through to the implementation stage. In this section we will take in an overview of these alternative tests and consider their strengths and weaknesses. Test items that were designed and subsequently abandoned will also be included here as illustrations of possible ways forward. In total, four alternative tests items were devised, two with subsidiaries. They will be referred to as the following:

1) 'Audio Response'
2) i) 'Silhouette' / ii) 'Picture Table'
3) i) 'Eyes-Closed' [Solo] ii) [Collaborative]
4) 'Personal Picture Description'

4.4.1 (1) 'Audio Response'
This item was designed to harness the accessibility of short sound samples in the pursuit of more culturally-valid test content. It is built on the assumption that simple audio clips could have a much wider cultural appeal than the pictorial ones currently preferred by Cambridge English. In keeping with the MDMR focus on this paper, candidates would have the opportunity to ‘experience’ easily identifiable audio test content and reflect upon, describe and speculate on what it brings to mind. The clips specified above are only suggested examples. Combined with further questioning this item is designed to be varied and adaptable.

Rationale: This item was designed to harness the accessibility of short sound samples in the pursuit of more culturally-valid test content. It is built on the assumption that simple audio clips could have a much wider cultural appeal than the pictorial ones currently preferred by Cambridge English. In keeping with the MDMR focus on this paper, candidates would have the opportunity to ‘experience’ easily identifiable audio test content and reflect upon, describe and speculate on what it brings to mind. The clips specified above are only suggested examples. Combined with further questioning this item is designed to be varied and adaptable.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Accessible, with greatly increased cultural scope and minimal room for western bias</td>
<td>• Implementation would necessarily be more complicated and ambient noise would need to be all but completely absent</td>
</tr>
<tr>
<td>• Potentially very affectively engaging and conducive to strong MDMR</td>
<td>• Some sounds may result in negative affective responses, thus jeopardising ethical clearance</td>
</tr>
<tr>
<td>• Flexible, with opportunities for adaptation</td>
<td>• Certain sounds could indicate cultural bias</td>
</tr>
</tbody>
</table>

Decision: Despite being conceptually very promising, the 'Audio Response' item was among those abandoned on ethical grounds. No guarantees could be made regarding the nature of the affective responses some sounds might elicit.

4.4.2 (i) 'Silhouette'
**Rationale:** Stylised, stripped-down imagery and cartoons have a long association with the Multidimensional Approach and ELT materials in general. The rationale behind this item was that such imagery (as used in the 'silhouette format') strips away detail and thereby both allows for greater MDMR and removes the cultural elements that potentially disadvantage some candidates. The example used here neatly shows how this can be done without eroding an item's affective qualities and the potential it has for speculation and description.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Reduced cultural bias through use of a more neutral and accessible 'silhouetted' image</td>
<td>• Some 'silhouetted' images, could conceivably cause distress and undermine ethical clearance</td>
</tr>
<tr>
<td>• Increased potential for creativity and the production of learner-authentic narrative(s)</td>
<td>• May require more specific subsidiary questions in order for B2 micro-skills to be encouraged</td>
</tr>
<tr>
<td>• Affectively engaging, depending on the content of the image used</td>
<td></td>
</tr>
</tbody>
</table>

**Decision:** This item received the same treatment as the 'Audio Response' one prior to it and was abandoned on precisely the same grounds (ethical considerations). The potential for negative affect was deemed too great to proceed without undermining ethical clearance.

4.4.2 (ii) 'Picture Table'
Rationale: This is effectively a more collaborative expansion of the previous item and many of the same comments and observations apply. Students would be asked to build a narrative around the pictures of their choice. The thinking behind this is that a 'Picture Table' grid could provide a culturally-open and flexible means of delivering test content that candidates can personalise and adapt to their own frame of reference.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Requires a high level of creativity and imagination</td>
<td>To be effective, this item may need to extend well beyond the requisite Cambridge test-item time-limit</td>
</tr>
<tr>
<td>Provides scope for MDMR and learner authentication</td>
<td>Some images may be deemed confusing, unsettling or culturally-biased in nature</td>
</tr>
<tr>
<td>Incorporates a rare kinaesthetic component . Highly flexible in terms of its content and the orientation of its subsidiary questions</td>
<td>As a relatively complicated and untraditional task, instructions may need a greater degree of clarification</td>
</tr>
</tbody>
</table>

Decision: This was the final test item to be rejected on ethical grounds and left unimplemented. Its time-frame, structure and the range of affect expressed were all considered problematic.

4.4.3 (i) 'Eyes-Closed' [Solo]
Rationale: To ameliorate the problems with the previous items, this was designed as a simple, clear and positively oriented ‘eyes-closed’ piece. The period of silent reflection and high degree of personal choice would meet this paper’s requirements for learner authentication and MDMR opportunities while also engaging affect.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Simple, short, clear and easy to implement</td>
<td>• Largely unstructured and may not be conducive to the production of B2 skill set</td>
</tr>
<tr>
<td>• High level of learner authentication and MDMR opportunities with minimal cultural bias</td>
<td>• Relatively easy for candidates to pre-plan and script their responses prior to the test.</td>
</tr>
<tr>
<td>• Potentially very affectively engaging</td>
<td></td>
</tr>
<tr>
<td>• Positively orientated, unlikely to cause distress</td>
<td></td>
</tr>
</tbody>
</table>

Decision: An item such as this, serving as a clear and direct test of MDMR’s potential in a test setting, was considered integral to this research project. The element of choice and open, positive orientation passed the non-malfeasance test and the item was subsequently administered and analysed.
4.4.3 (ii) 'Eyes-Closed' [Collaborative]

Rationale: Alternative test item five was a collaborative extension of the previous 'eye-closed' item in which candidates would have the opportunity to respond to their partner's work (after a suitable period of self-expression). This could take the form of a question, discussion or comment, further questioning then being guided by the examiner.

Advantages
- Relatively clear and uncomplicated item
- Potentially affectively engaging and interactive with learner authentication and MDMR opportunities
- Reduced levels of cultural bias

Disadvantages
- Potential disconnection between candidates – [relies on the coherence and accessibility of Candidate A's initial response]
- As above, largely unstructured and may not encourage use of the B2 micro-skills

Decision: Due to time constraints, this item was not administered. The merits of a more collaborative MDMR approach were instead investigated through the subsequent 'Personal Picture Description'. It was thought particularly appropriate to trial MDMR in an 'interactive' test setting so as to better judge the effects it had on output and discourse management.
4.4.4 'Personal Picture Description'

Rationale: The basis for this item was to turn the photographic format used in the FCE towards more personal ends. Candidates would necessarily have a well-established, authentic relationship with the test content and would be invited to reflect on it in a manner conducive to the formation of MDMR. It is also envisaged that such an item could be expanded to include an interactive component in which candidates question and comment on each other’s pictures and the 'representations' that follow.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Highly learner-authentic in its content</td>
<td>• Very easy for candidates to pre-plan and script their responses prior to the test.</td>
</tr>
<tr>
<td>• Provides opportunities for MDMR</td>
<td>• Detailed visual stimuli may actually inhibit strong Multidimensional Mental Representation</td>
</tr>
<tr>
<td>• Positive and affectively engaging</td>
<td></td>
</tr>
<tr>
<td>• Absolutely minimal levels of cultural bias</td>
<td></td>
</tr>
<tr>
<td>• Simple, short, clear and easy to implement</td>
<td></td>
</tr>
</tbody>
</table>

Decision: 'Personal Picture Description' was also considered ethically sound and suitably learner authentic for its use in the latter stage of this research. As with item 4.4.3 (i), its results will be discussed further on in this analysis.

Having looked over the suggested alternatives, we now turn to the first FCE item undertaken ('Museums'). This will act as a base line against which to judge the its alternative equivalent, namely item 4.4.3.(i)
4.5 FCE Test Item 1

'Museums' (Focus: Candidate B)

Interlocutor: In this part of the test, I'm going to give each of you two photographs. I'd like you to talk about your photographs on your own for about a minute.

(Candidate B), it's your turn first. Here are your photographs. They show people visiting different museums.

Place Part 2 booklet, open at Task 2, in front of Candidate B.

I'd like you to compare the photographs, and say what the people might find interesting about visiting these different museums.

All right?

Candidate B: ..............................................................

Interlocutor: Thank you. (Can I have the booklet, please?) Retrieve Part 2 booklet.

What might the people find interesting about visiting these different museums?
FCE Test Item 1: Candidate B Performance

Having warmed up with (and fared relatively well in) the 'Personal Questions' item (Appendix B), Candidate B was then asked to answer a question taken from the picture-based original 'Long Turn' section of the FCE paper (Item 2). The transcript of his response is detailed below:

4.5.1 'Museums' Transcript

<table>
<thead>
<tr>
<th>Examiner:</th>
<th>[instructions given]...</th>
<th>[directed at Candidate B] So...can we start here?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Candidate B:</td>
<td><em>Errr.... It's difficult to tell you because I have never never visit a museum.</em></td>
<td></td>
</tr>
<tr>
<td>Examiner:</td>
<td>Ah.</td>
<td></td>
</tr>
<tr>
<td>Candidate B:</td>
<td><em>I'm not interested in that</em></td>
<td></td>
</tr>
<tr>
<td>Examiner:</td>
<td>That's okay. No problem at all.</td>
<td></td>
</tr>
<tr>
<td>Candidate B:</td>
<td>Depends on the person, you know. It's like.. it's interesting to some people. Maybe it's art or. <em>I don't know.</em></td>
<td></td>
</tr>
<tr>
<td>Examiner:</td>
<td>Okay.. You do not like..</td>
<td></td>
</tr>
<tr>
<td>Candidate B:</td>
<td>Ummm, <em>I'm not interested in that</em>. I'm 'trested in gyms, sports.. these kinds activities.</td>
<td></td>
</tr>
<tr>
<td>Examiner:</td>
<td>Okay.. okay.. So why <em>might</em> someone like those museums?</td>
<td></td>
</tr>
<tr>
<td>Candidate B:</td>
<td>Ummm, maybe someone 'trested in dinosaurs? .. to see the [names?] and the [bones]? Ummm.. or he do research – like you?</td>
<td></td>
</tr>
<tr>
<td>Examiner:</td>
<td>Maybe.</td>
<td></td>
</tr>
<tr>
<td>Candidate B:</td>
<td>Kind of..</td>
<td></td>
</tr>
<tr>
<td>Examiner:</td>
<td>Maybe in History? Okay.. yeah.</td>
<td></td>
</tr>
<tr>
<td>Candidate B:</td>
<td>[pointing to the gallery picture]: Maybe he has a picture of like.. which explains the famous peoples of like actors, presidents..</td>
<td></td>
</tr>
<tr>
<td>Examiner:</td>
<td>Maybe.. Like photographs, or art?</td>
<td></td>
</tr>
<tr>
<td>Candidate B:</td>
<td><em>Unmm.. [shakes his head]</em></td>
<td></td>
</tr>
<tr>
<td>Examiner:</td>
<td>No, that's fine.</td>
<td></td>
</tr>
</tbody>
</table>

Observations:

4.5.2 Content Engagement

As is clearly evidenced, Candidate B struggled to engage with this item. Whereas he could identify the pictures as being of 'Museums', they are opening stated as such in the paper itself. Consequently, no credit can be given for this. The actual experience of attending a museum, experience that would assist him greatly in successfully answering the question at hand, is fundamentally alien to him ('It's difficult to tell you because I never never visit a museum'). Even with gentle prompting redirecting him towards the content of the question ('Why *might* someone like those museums?') he fails to produce any responses likely to earn him a pass. As the candidate repeatedly states ('I'm not 'trested in that'), affective engagement is all but absent. Beyond having no frame of reference for the subject matter, he also has no interest in the prospect of engaging with it.
4.5.3 Grammar and Vocabulary

The candidate never breaks out of using the Present Simple tense, even when trying to express other simple grammatical forms like the Present Progressive (Maybe he do [is doing] research – like you?). This all but guarantees failure in any intermediate examination. Albeit extremely limited, there is some intermediate-level vocabulary range in evidence (‘which explains..’, ‘depends on the person..’), indicating that the candidate may have sufficient lexis to produce better responses in relation to more appropriate subject matter. Perhaps in part due to the aforementioned lack of engagement, the candidate simply does not produce enough output for his vocabulary to be accurately assessed. It is to this point we now turn.

4.5.4 Output, Pronunciation and Fluency

Remarkably, the candidate took well over a minute to produce the above, resulting in an output level of only 40 words per minute. His lack of a meaningful connection to the item’s content is born out not only in the quality of his output, but its quantity as well. This is in sharp contrast to the ‘Personal Questions’ item preceding it, which was not picture-based and in which average output between candidates was recorded at 113 words per minute.

Pronunciation in the ‘Museum’ item was communicative but generally sub-standard, containing some unintelligible utterances, unnecessary abbreviations and dropped syllables. So too with fluency, did the candidate struggle, leaving protracted periods of silence and never really communicating with full, consecutive and coherent sentences. There is some frustration and confusion evident in both his intonation and body language, and this seems to work against him getting into any kind of a rhythm.

Whether this poor overall performance is attributable to the weakness of the candidate or deficiencies in the test itself will be subject to discussion later in this paper.

4.5.5 'Museums' Marking: Candidate B

<table>
<thead>
<tr>
<th>Grammar &amp; Vocabulary</th>
<th>Pronunciation</th>
<th>Discourse Management</th>
<th>Interactive Communication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mark [-] 1-5[+]</td>
<td>1</td>
<td>2</td>
<td>&lt; 2 [-]</td>
</tr>
<tr>
<td>Rubric</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;Shows a degree of control of simple grammar. Candidate uses some appropriate vocabulary when talking about everyday situations&quot;</td>
<td>&quot;Is mostly intelligible. Sentence and word stress is generally accurately placed. Individual sounds are generally articulated clearly&quot;</td>
<td>&quot;Produces responses which are extended beyond short phrases, despite hesitation. Some repetition. Uses basic cohesive devices&quot;</td>
<td>&quot;Limited appropriate response. Struggles to keep interaction going even with prompting and support&quot;</td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Global Achievement Grade [Overall Score]: 5.5: Borderline: A2/B1 28% [Fail]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4.6 Solo Eyes-Closed Alternative Item (Focus: Candidate B)

4.6.1 'Solo Eyes-Closed' Transcript

| Examiner: | [period of prolonged silence] Can you? Anywhere.. anywhere you feel most comfortable. |
| Candidate B: | My father's farm, I've got like a farm in Kuwait.. mostly, like, every Kuwaiti has a farm.. Once I go at night.. quite late.. no one there.. and go to touch my horse. He's called 'Zeirun'. Stay with him. Sometime watch him.. Clean. Ride the horse for one hour.. two hour. No one there. Alone... Don't think about anything.. just riding the horse. Sometimes I talk to him. It's my best time. |
| Examiner: | Yeah, tell me about the horse.. I want to 'see' the horse again. |
| Candidate B: | I treat the horse by different ways. I treat the horse like a 'guest'. It's like a guest, because the horse I think like someone you all the time care about. Like, one month, I didn't come to my farm to see my horse. He's all the time like - the guys that work there - all the time he's just 'shouting', 'shouting' and when I came, he calmed down. Just what I'm like. No one can ride him. No one can ride him. No one.. |
| Examiner: | Except you? |
| Candidate B: | Except me. |
| Examiner: | What can you.. what can you hear at night? |
| Candidate B: | Just 'air' |
| Examiner: | Just air; nothing? |
| Candidate B: | ..and the sound from the horse. It's really amazing.. amazing! |

Observations:

4.6.2 Content Engagement

The content of this item, being entirely generated by the candidate, stands out as being personal to him and the culture barrier appears to be all but completely absent. Candidate B went on to describe this as the most enjoyable item he took part in and evidence for this heightened level of affective engagement is stark in comparison to the previous item ('It's my best time', 'It's really amazing.. amazing!', 'Someone you all the time care about').

His connection with the content generated is born out in the interactivity and 'physicality' of the language he uses (verbs such as 'touch', 'ride' and 'clean') and this was often accompanied by hand gestures and facial expressions while eyes were closed. The closeness of his connection to the item's content is played out, at least in part, in a physical way during the re-experiencing process. These results do appear to map onto expectations of how multidimensional imaging would perform in such a setting.

It is also interesting to note that, without prompting, the Candidate reflects upon his own mental processes (or lack of them) while immersed in his representation ('Don't think about anything'). This can be seen as a further possible indicator of deep connection and engagement.

4.6.3 Grammar and Vocabulary

The MDMR eyes-closed item appeared to have little substantive impact in this area. Whereas Candidate B continues to not use complex forms, there are some very minor improvements evident. Present Simple still abounds though the candidate does recourse to the Past Simple on occasion 'I didn't come to my farm', 'I came', 'he calmed down'), though basic errors are also apparent ('Once I go [went] at night'). Similarly,
Present Progressive is also employed ('Riding the horse', 'He's just shouting') and this alone marks his response out as being more grammatically diverse than his previous effort. Otherwise valuable instances of the Present Perfect (I've got) are basic and more indicative of the idiomatic form synonymous with simply 'I have'. Consequently, little credit can be given here as no other examples are evident. The vocabulary used is generally satisfactory and fit for purpose but far from indicative of a strong performance. No significant improvements are evident in this regard, though the candidate's use of the 'guest' metaphor ('I treat the horse like a guest') is creditworthy and conforms to the requisite skill-set.

4.6.4 Output, Pronunciation and Fluency

Overall, output improved greatly. There was no confusion regarding the content or what the candidate was required to do and, as a result, no portion of the test was devoted to clarification. In terms of quantity alone, output rose to around 110 words per minute and this marks a significant improvement on Candidate B's previous efforts. The item drew to a close because the time limit had been reached but the candidate gave every indication that he'd be happy for questioning to continue. Interestingly, pronunciation and fluency improved as the item went on. The first response reads as if it was a 'stream of consciousness', with regular pauses, short utterances (sometimes reduced to single words as in 'Alone.' and 'Clean...') and some unintelligibility. This was not picked up on as a possible drawback during the design phase, though perhaps should have been. The MDMR process, initially at least, seems to break up speech while the candidate acclimatises to the 're-experiencing' process and his 'new' surroundings. While he 'sets the scene', Candidate B employs stripped-down language to 'flesh out' his mental representation while he is experiencing it. He is not trying to impart information about a static image – he is involved in a process of 'noticing' and describing an experience that is essentially alive and fluid. This has problematic implications for the test format, to which we will return.

However, fluency does improve markedly once the 'stage is set'. The pauses are all but absent, the candidate's rate of speech increases and he seems more inclined to speak in fuller sentences. The reasons for this are unknown. We could speculate that it is simply a matter of a reduction in anxiety or inhibition as the test progresses and the candidate finds his feet. Equally, the candidate is essentially describing a scene or situation in his first response, whereas he describes his relationship with his horse in the second. The latter arguably represents a more static construct than the former, and therefore may be more easy for him to give form to.

There is some repetition in this item, though it is often a specific form of repetition that serves an affective or communicative function. If we consider; 'it's amazing... amazing!' or even 'shouting, shouting', we can appreciate that they do not necessarily show the candidate struggling for words but in fact represent an effort to accentuate certain aspects of his message.

4.6.5 'Eyes-Closed' [Solo] Marking: Candidate B
4.7 FCE Test Item 2:

'School Activities' (Focus: Candidates A & B)

Overall Global Achievement Grade [Overall Score]: 11: Borderline: B1/B2 55% [High Fail]
Candidate A & B Performance

4.7.1 'School Activities' Transcript

Examiner: Now I'm going to show you two photos to discuss together, if possible.

There are these photos. Actually, there are seven photos here.

Candidate B: Interesting

Examiner: Just take a moment to look at those photos...

Here are some different activities that students often do during the school day.

[Candidates take some time to look at the photographs]

So, why might it be important to do these activities at school?

Candidate A: Errr, I think because, to discover like, each person has ability. errr.. skills in a part of life, many activities and to discover this errr; they make student try all kind of as possible. Here, computers, sport... I think this like reading or sing poem or something like that, cooking and these traditional things, and, to discover like [unclear – everything?] Or to give them a bit also like skills, like they might need it in future or something like that.

Examiner: [directed at Candidate B]: Okay, do you agree?

Candidate B: I agree with my friend. I want to add something. Normally at school we start from the morning and afternoon like six or seven hours. So we can't say like six or seven hours all the students want to study. They should have like some free classes.. some activities to do. Like some people they love video games, they can't have.. classrooms, computer rooms can't spend that. Also, there is some people, whatever, girls or female or male or female they love to cook. Like when I was in high school they have, there is a free class, you have like music to record or you wanna learn something there is like two or three teachers, they help us. also there is free like, between classes there is a free time, like a break.

Here they eat [pointing at eating photo] and here maybe they 'song'.

Examiner: Maybe they are singing. Yes.

-----------------------------------------------

Examiner: Which two activities do you think are not so important?

Candidate A: Not important?.. I think singing and dancing.

Candidate B: Actually, all of them I think is important. I can't say anything is so important, but that one I... I don't have any idea about, what's that actually? [shrugs shoulders] So I can't say if it's important or not.

Examiner: I think maybe it's acting or theatre.. or 'drama'.

Candidate B: It might be, but in my culture, in my countries – we never ever have it.

Examiner: ..and what was your favourite subject at school? Always sport?

Candidate B: Always

Candidate A: I like.. well like, errr 'Math', especially the teacher like. The teacher makes you like the subject. You will be good and you will love the subject.

Observations:

4.7.2 Content Engagement

As evidenced in both the candidate's 'background perspectives' and the latter stages of the transcript, there were areas of this item that made comment difficult ('Drama Picture' [top right] and 'Cooking Class' [middle]). However, despite this narrowed scope the remaining pictures did generally reflect their own
school experience and provided a satisfactory basis for conversation. There is very little in the way of genuine and positive affect, though some limited degree of engagement is found in Candidate A’s final comment and Candidate B’s first. The latter later openly voices his frustration regarding some of the item’s content, though this negative affect does not prevent him commenting at some length on the other pictures. Whereas their focus of the question is somewhat lacking, they do at least seem cognitively engaged in the process of answering it and provide some interesting insights.

4.7.3 Grammar and Vocabulary
Both candidates’ grammar was considerably sub-par. Candidate B again failed to break out of using the Present Simple tense and much the same can be said for his partner (with the exception of Future with ’Will’ [You will be good and you will love the subject]). Candidate A also employs a number of gerunds (’reading’, ’cooking’, singing’, ’dancing’), though never uses the ’-ing’ form to express the Present Progressive despite numerous opportunities to do so. Candidate B also fails to take advantage of these (’Here they eat’ not ’Here they are eating’).

In response to the question (’Why might it be important..?’), candidates do make efforts to speculate with their language, (A: ’They might need it in future’, B: ’Maybe they.’) which elevated both of their scores fractionally. Candidate B occasionally accompanies these with other modals such as ’can’ (’We can’t say’) and ’should’ (’They should have some free classes’), but overall they are not enough to deliver him from a deficient overall mark. The vocabulary of both candidates is relatively basic but generally appropriate to the task at hand. Were the grammar to have been at a higher level, it may have been sufficient to sustain a passing grade.

4.7.4 Output, Pronunciation and Fluency
The overall quantity of output was very satisfactory, with an average of just over 100 words per minute. The division of speech-time was in Candidate B’s favour, though his partner’s contribution was more than sufficient. This represents a great improvement on what had gone previously. However, each can be seen to differ in their levels of pronunciation and overall fluency. Despite some incoherence, Candidate B generally speaks in full sentences and puts in a more assured performance. He also displays some strong discourse management that links the exchange together well (’Actually..’, ’I agree with my friend’, and ’I want to add something’). This is in contrast with his partner who gives some unintelligible utterances and at times struggles to get a coherent message across. This, combined with higher levels of hesitation, contributes to a mark three points lower than that of his partner. Both candidates heavily overuse the word ’like’, which is almost invariably employed inappropriately as a kind of weak connective or space-filler. However, Candidate B performs better overall as can be seen in the marks overleaf.
### 4.7.5 'School Activities' Marking:

#### Candidate A

<table>
<thead>
<tr>
<th>Category</th>
<th>Grammar &amp; Vocabulary</th>
<th>Pronunciation</th>
<th>Discourse Management</th>
<th>Interactive Communication</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mark</strong> [-1–5[+]</td>
<td>2</td>
<td>&gt; 1 [+</td>
<td>&gt; 1 [+</td>
<td>2</td>
</tr>
<tr>
<td><strong>Rubric</strong></td>
<td>&quot;Shows a degree of control of simple grammar. Candidate uses some appropriate vocabulary when talking about everyday situations&quot;</td>
<td>&quot;Is mostly intelligible. Sentence and word stress is generally accurately placed. Individual sounds are not always articulated clearly&quot;</td>
<td>&quot;Produces responses which extend well beyond short phrases, despite hesitation. Uses only basic cohesive devices&quot;</td>
<td>&quot;Generally responds appropriately. Little effort to develop interaction or comment on partner response&quot;</td>
</tr>
</tbody>
</table>

**Overall**

Global Achievement Grade [Overall Score]: **7.0**: B1   35% [Fail]

#### Candidate B

<table>
<thead>
<tr>
<th>Category</th>
<th>Grammar &amp; Vocabulary</th>
<th>Pronunciation</th>
<th>Discourse Management</th>
<th>Interactive Communication</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mark</strong> [-1–5[+]</td>
<td>2</td>
<td>&gt; 2 [+</td>
<td>3</td>
<td>&lt; 3 [-]</td>
</tr>
<tr>
<td><strong>Rubric</strong></td>
<td>&quot;Shows a degree of control of simple grammar. Candidate uses some appropriate vocabulary when talking about everyday situations&quot;</td>
<td>&quot;Is intelligible. Intonation is generally appropriate. Sentence and word stress is generally accurately placed. Individual sounds, generally articulated clearly&quot;</td>
<td>&quot;Produces extended stretches of language despite some hesitation. Contributions are relevant and there is very little repetition. Uses a range of cohesive devices.&quot;</td>
<td>&quot;Initiates and responds appropriately. Maintains and develops the interaction and negotiates outcome with little support.&quot;</td>
</tr>
</tbody>
</table>

**Overall**

Global Achievement Grade [Overall Score]: **10**: B1   50% [Fail]

---

### 4.8 'Personal Picture Description' (Focus: Candidate A [with B as contributor])

In line with Alternative Item 4.4.4 (p.40), the candidate shows the examiner a picture and then reflects on it in silence with eyes closed. Candidate B is also shown the picture prior to reflection. The picture Candidate A has chosen in this instance depicts a group of five men sitting and kneeling in a carpeted room. They are of various ages and mostly wear traditional Arab attire. It cannot be reproduced here as this would violate ethical consent and the terms agreed with the candidates prior to the item's trial (i.e. that only recordings and short-hand notes would be taken from the interviews and tests).

The results of this item are transcribed overleaf.

---

49
4.8.1 'Personal Picture Description' Transcript

[period of silence]

Candidate A: Ah.. I think it's the best time, like errr.. or the best favourite place where we eating like errr.. in my home.. We sit, me and my brothers.. my sisters and errr.. my brothers wives.. their children.. my father, mother and we eat and errr.. Before we start eating we have.. you know, like 'argument' and laugh together and errr and have a nice time and when, actually during the eat.. we also eat and errr speak. After that we don't move like, we keep speaking and laughing and together and telling each other what happened this day and I think this is the best moment like in my life.

Examiner: What can you.. what would you eat?
Candidate B: It's rice with chicken, rice with lamb.
Candidate A: Yeah, it's with yellow rice and chicken.
Candidate B: It's from Yemen. Every.. all the rice, all the chicken, of course, is coming from Yemen..
Examiner: Yes?
Candidate B: ...it's become a tradition for us as well.
Examiner: Okay
Candidate B: ... because we're used to it.
Candidate A: Yeah, and we like errr.. it depends on the food because every time we go there is bread, rice – white, yellow and err.. every day.
Candidate B: Every day for lunch, we go have rice.
Candidate A: And this I think.. we call it 'Durahiya'. It's a place where man sit.
Candidate B: Yeah, I love it!
Examiner: That's good
Candidate B: Fantastic!
Candidate A: Sometimes fish.. sometimes fish, sometimes hamburger [unintelligible]. The best thing.. we sit together and we speak and that time not allow us to use our phones.. you know, and I think that's the best time or the best.. favourite place.
Candidate B: It's like a private place with people you love... really love.
Examiner: Of course.
Candidate B: ...and there's like, a relationship with everyone. Like.. 'love'. You don't think like you 'have a job' or anything.. just.. but we don't care about outside because we sit inside the living room and there is air conditioning everywhere.
Candidate A: This my brother.. this like is very common place my home. All the time spend[?] most my time there. Can sit there, eat there, sleep there.. Every.. All the time. During I sit at home I sit there. I think it's the most comfortable place for me and my brothers also.
Examiner: So what are they doing here now? What *might* they be doing? Are they just talking?
Candidate A: I think that now they are fasting. Just sit and read and, and then prayer. Some stuff like that, but in this space it's like all guys they sit there and speak and they can't eat right now (Ramadan). They watch TV, they speak. If they want to sleep also, they sleep there. They just bring cover and pillow. I think that very common place. They all the time sit there and sometimes, like my friends, they sleep in my home like in this space we'll be together – all the time together. Because we don't have responsibility as we will have like in future.. married, childr.. and we enjoy our time there. Yeah, everything there – everything, everything! If we have problem we sit there, it's the best place.
Examiner: Okay, perfect

[Ends 3:02]
**Observations:**

### 4.8.2 Content Engagement

The enthusiasm both candidates felt for this item was palpable and the transcript above is littered with comments that indicate a high level of affective engagement (Candidate A: 'the best moment in my life', 'favourite place', 'all the time together' Candidate B: 'I love it!', 'fantastic!', 'a private place with people you love... really love'). The scenes described are very warm and personal in nature. They are of a place where problems are solved, where family and friends are close by and where the outside world is irrelevant ('we don't care about outside'). Both candidates seemed truly immersed in the experience and neither struggled with what they were instructed to do.

Interestingly however, the 'physicality' of the language evidenced in Candidate B’s earlier MDMR response is all but completely absent here. So too, are the gestures and facial expressions that accompanied his solo response greatly reduced. These missing outward signs of internal mental representation are somewhat puzzling, though possible explanations can be found in the associated literature. As previously discussed, the presence of the picture itself may distract from, or substitute the need for, deeper MDMR. Whereas this item certainly conforms to its learner-authentic design, in doing so it may have reduced its potential for MDMR.

Looking back at the content generated in this item, we can detect a problem with it in that both candidates are from the same cultural background and can easily relate to what each other say. The results may well differ if the two candidates were selected from disparate cultures, though equally this could provide fertile ground for inter-candidate questioning and speculation (something that is all but absent from the above).

### 4.8.3 Grammar and Vocabulary

Grammar and vocabulary are again at a very similar level to the previous items. Given the nature of the item itself, Candidate A's greater output leads to him producing the bulk of the (limited) grammatical diversity present. Here we find an abundance of Present Simple and some instances of the Present Progressive ('eating', 'laughing', 'fasting'). Alongside there is a use of the Future with 'Will' ('...as we will have [...] in future') and a First Conditional ('If we have a problem, we sit there'). Additionally, Candidate A shows a limited degree of command over simple speculative language when he responds to 'What might they be doing?' with 'I think they are fasting'. However, these forms are basic and insufficient to score high marks at B2 level, though it is worth noting that the more complex of them are again weighted towards the end of the piece.

Despite some indications to the contrary ('it's a *traditional* place', 'a *private* place', 'we don't have responsibility') both candidates generally show their vocabulary to be lower than the requisite B2 standard.
What stands out about this piece is the sheer volume of output produced. Together the candidates speak at close to 170 words per minute. Given the design of the item, this was naturally heavily in favour of Candidate A, who had previously been the less confident of the pair. This sharp increase in output was unexpected, particularly with such minimal prompting. It is difficult to determine whether the three-way ‘conversational’ orientation of the piece was the driving force behind this increase, or whether it was more attributable to the affective and personal nature of the content. However, given that the results are much higher than either the original FCE items or the ‘Solo Eyes-Closed’ alternative, we might speculate that both content and design played their parts.

As has become a recurring theme throughout these items, pronunciation shows only slight variation. It seems difficult to alter such a fundamental and engrained aspect of speech through test-design alone.

Reflection and engagement time had no discernible impact on word formation.

Just a cursory glance at the first and final responses (both from Candidate A) shows similarities with the Candidate B’s previous MDMR item. There is an early period of acclimatisation, marked by multiple hesitations and pauses. This sporadic, uneven output seems to slowly dissipate at the item continues and fluency does improve, albeit from a relatively low base. Some of his best work of all the items can be found in this final paragraph as he reflects on his home, his friends and the responsibilities the future holds.

Admittedly, there is a problem with repetition throughout, with ‘eat’ (seven iterations) and ‘speak’ (five iterations) being prominent examples. Candidate A in particular is guilty of this, though both candidates maintain their overuse of the word ‘like’ (fourteen iterations). Interestingly, both Candidates A and B again seem to occasionally use repetition enthusiastically for deliberate emphasis or emotional resonance. We see this at work with Candidate B’s ‘a private place with people you knew, really love’ and Candidate A’s ‘Yeah, everything there – everything, everything!’ It seems unjustifiable to mark them down for these affective repetitions, though the Cambridge rubrics do not allow for this.

### 4.8.5 ‘Personal Picture Description’ Marking: Candidate A

<table>
<thead>
<tr>
<th>Category</th>
<th>Grammar &amp; Vocabulary</th>
<th>Pronunciation</th>
<th>Discourse Management</th>
<th>Interactive Communication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mark</td>
<td>&gt; 2 [+</td>
<td>2</td>
<td>2</td>
<td>&lt; 3 [-]</td>
</tr>
<tr>
<td>Rubric</td>
<td>&quot;Shows control over simple grammar and attempts some complex forms. Uses a limited range of appropriate vocabulary&quot;</td>
<td>&quot;Is mostly intelligible. Sentence and word stress is generally accurately placed. Individual sounds are generally articulated clearly&quot;</td>
<td>&quot;Produces responses which are extended well beyond short phrases, despite hesitation. Some repetition. Uses basic cohesive devices&quot;</td>
<td>&quot;Initiates and responds appropriately. Maintains and develops the interaction and negotiates outcome with little support.&quot;</td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td></td>
<td></td>
<td>Global Achievement Grade [Overall Score]: 9: B1 45% [Fail]</td>
</tr>
</tbody>
</table>
Chapter 5: Conclusion

By way of conclusion, it is perhaps best to return to the main questions that have guided this research. The following sections will therefore focus on the cultural orientation of the FCE’s visuals and the evidence for MDMR’s testing potential. These points will be addressed in a critical and reflexive way, with a view to continuing and improving this research. In closing, recommendations will be made for Cambridge English’s ongoing efforts in the testing field.

Do FCE visuals exhibit Cultural Bias?

It must be conceded at the outset of this section that the small number of candidate perspectives feeding into this research was insufficient to provide conclusive data of statistically significance. That having been said, the input from both candidates did support many of the contentions of this paper. They had problems relating to and even recognising some of the scenes depicted, and this proved a stumbling block in terms of test performance, rather than an opportunity to speculate and enquire. Both candidates also acknowledged that the pictures were almost exclusively rooted in the experiences of specifically ‘English’ or ‘American’ people. The few clear exceptions to this, such as ‘Tribal Music’, were cultural extremes that are not in-keeping with the examination’s ‘real-life situations’ mandate. Based on the limited evidence gathered, we can conclude that there are at least some indications of unnecessarily ‘encultured’ subject matter. Clearly interviewing a wider and more culturally-diverse pool of candidates would be necessary to add more weight to this claim.

Looking back over the picture portfolio, any claims that the representations therein are ‘fair and relevant to candidates of all backgrounds’ seem difficult to justify. The boldness of these claims and the extent to which the FCE falls short of them is all the more peculiar given the rigour of Cambridge’s research in this area. The Broadening of Cultural Context Initiative and its resultant recommendations make a very strong case for greater awareness of the cultural orientation of test content. This has implications ranging from how such examinations are designed through to how the backgrounds, experiences and motivations of its candidates are perceived. The reasons for the Initiative’s findings not having yet fully permeated Cambridge’s examinations remain unclear. Whereas some of the pictures used in this paper date back to 2008, many of the items actually administered were much more recent, with some having been published in 2015. It seems unlikely that the natural delay in putting research into practice would take as long as this. Coupled with the very low number of citations associated with the Initiative, we are left to assume that it played little or no role in shaping Cambridge English’s examination content at this level. This point will be returned to in the following recommendations (p.56).
To what extent can Multi-Dimensional Mental Representation help neutralise this bias?

Despite the necessary ethical constraints placed upon its usage, MDMR has yielded interesting results throughout this research. Prominent among these were output and affective language, both of which rose sharply when compared with the more traditional items. To witness the candidates speaking so freely and openly about the things that they care about most was both a highly desirable research outcome and a personal privilege. MDMR did seem to deliver on its potential to open up expressive spaces through which candidates could generate their own content. Furthermore, no evidence for any significant culture barrier to communication was evident at any stage of these exchanges.

Whereas both candidates did produce overall scores higher in the MDMR items than those based on the original FCE pictures, these scores were still lower than expected when taking into consideration their IELTS level. The improvements noted were possibly a natural result of greatly increased output and, based on the limited evidence available, it cannot necessarily be assumed that they are directly attributable to MDMR. A stronger case can be made for the impact MDMR has on affective engagement and output, making MDMR a possible indirect cause of these improvements in overall test performance. The candidates themselves did not deliver at a level appropriate to their previous B2 credentials, though this too could be a result of factors other than proficiency, such as unreadiness or the fact the tests took place during a period of fasting. Were further research in this area to be carried out, it would be advisable to conduct pre-tests to ascertain the level of its participants and not take previous scores at face-value. Unfortunately, only one cycle of tests and interviews was possible in the time-frame of this particular paper.

Whereas there were indications that MDMR could provide a possible ‘way forward’ in the testing field, a number of caveats need to be attached to this claim. The ‘slow start’ problem, in which candidates’ speech becomes disjointed during the early stages of MDMR, is potentially a considerable drawback. Further trialling would be needed to determine whether this is a natural result of the MDMR process or if it is attributable to more general inhibiting factors (anxiety, or cognitive unreadiness). It would be worthwhile exploring whether this phenomenon persists in native speakers while engaged in the MDMR process. Doing so may help determine how fundamental a problem this to L2 learners. Solutions could take the form of initiation exercises or deliberately only assessing the latter stages of the output, though admittedly both are undesirable concessions for a test of this scale to make.
A further concern relates to the design of the tests themselves. Given that the MDMR items were essentially first drafts, they did not allow sufficient opportunity for candidates to fulfil Cambridge's rubrics, nor were they satisfactorily orientated towards the execution of B2 micro-skills. If time had allowed for a second or third round of tests it would have been fruitful to tailor the follow-up questions towards these ends and thereby encourage detail, speculation, better discourse management and more complex grammar.

While learner-authentic and all but devoid of cultural bias, the content generated by the candidates also presents us with some problems. If administered in the manner outlined in the early stages of this analysis, there exists a great deal of opportunity for candidates to script their responses prior to the test taking place. It is hoped that this too can be ameliorated by making the items more unpredictable by presenting candidates with a more specific initial MDMR 'setting', or by introducing a wider range of subsidiary questioning from which the examiner can draw.

Another more serious concern relates to the assumed universal viability of experiential learning across candidates and, indeed, across cultures. It is likely that students from some backgrounds may be very unfamiliar with experiential approaches and culturally predisposed towards more mechanistic, detached or 'transmitive' forms of learning (and thus examining also). Such unfamiliarity could nullify any potential increases in personal relevance, output and engagement alternative items may offer, again showing the need for caution and further research in this area.

Finally, it bears repeating that this paper has not been able to fully explore the range of test-items that could have been based around MDMR. Some of the items abandoned on ethical grounds contained audio and kinaesthetic components which could have provided highly valuable indicators of MDMR's wider testing potential. As things stand, we only have a snapshot of this potential from which to draw conclusions, and these conclusions are necessarily limited as a result.

Despite its notable strengths, MDMR cannot on its own be considered a solution to cultural bias in international assessment at such an early stage of its research. Its use in a testing environment does however, draw attention to a number of idiosyncrasies in its application that, in themselves, are welcome findings for this paper to have uncovered. The results gleaned at least establish that MDMR does warrant further investigation as an alternative testing tool, particularly in the areas of output and affective engagement.
5.1 Closing Recommendations

Reflect on the choice of representations used in the FCE and similar examinations

If visual representations are deemed absolutely necessary for the examination's format (and it is by no means clear that they are), it is incumbent upon Cambridge to exercise greater discretion and balance in its choice of material. Simply committing to the portrayal of a wider selection of ethnicities sidesteps a range of deeper issues around cultural bias (context, activity, gender roles). Alternative approaches could take the form of region-specific versions of the FCE, or simply a global paper that was less western-orientated. While it is accepted that region-specific papers would undermine test standardisation, it should be noted that Cambridge have already shown a degree of flexibility towards subgroups of FCE candidates through their introduction of the FCE for Schools.

Incorporating a wider array of choice for examiners to tailor content to specific cultural contexts may also yield positive results, though appealing to examiner discretion may be problematic (See above, Research Limitations p.27) and the size of a greatly expanded bank of images could prove unmanageable. A simple compromise that would not rely on either the cultural discretion of the examiner or the issuing of region-specific papers could come in the form of devolving decision-making to the candidate(s). Offering them a balanced selection of images from which they can choose could go a long way to improving cultural validity.

If a smaller selection of suitable visuals was required, this could be achieved through more neutral approaches to representation such as silhouettes, cartoons, abstract imagery and pictures that may not even portray people as subject matter at all.

Better understand the global nature and needs of your audience

Candidates need not be seen as a homogeneous group of people external to western culture who hold aspirations of better engaging with it. Rather, they are individuals with their own diverse motivations for a relationship with the English Language. These motivations, as Cambridge themselves rightly note, may not be driven by a desire to even learn about western culture, let alone assimilate into it. The FCE is currently marketed as an upper intermediate gateway exam that has a range of social, professional and academic purposes for its candidates. Some are pitched as being specific to living and working in an English-speaking environment, while others are left more general and undefined. It may be advisable to offer alternative tests with different aspects of this these broad purposes in mind, rather than include them all in the same paper. Allowances have already been made for the needs and interests of school-aged learners through the introduction of 'FCE for Schools'. It could be equally viable to offer 'Professional', 'Academic' and 'Social' packages that have a greater or lesser degree of cultural specificity depending on their purpose.
Be more open about the current cultural orientation of the FCE

In the event that change on this front is not forthcoming, Cambridge must at least be more up-front with its audience regarding the western-centric content of its examinations and the reasons behind it. Phrases from its candidate/teacher literature such as ‘real life situations’ are vague and beg a host of questions. It is recommended that Cambridge engage with, in detail, the questions that have been levelled at its tests for decades, specifically: to what extent does the FCE aspire to be a ‘purely linguistic-type’ test (Joint Committee: 1962 p.17) and on what grounds can its cultural content be justified in the current climate?

Re-evaluate the rubrics of your speaking papers

Certain aspects of everyday speech, such as hesitation and repetition for emotive emphasis, are generally penalised by Cambridge’s speaking rubrics. Yet on the evidence here they can, at times, be seen as a natural product of enthusiasm, excitement and deep affective engagement. This is not to say that all instances of hesitation and repetition should go unchallenged, but that some have an important and deliberate communicative function.

Build on and implement the research you have already carried out in this area

The background research conducted by the Broadening of Cultural Context Initiative provides an excellent base from which to re-evaluate Cambridge’s approach to test content. Its recommendations are unequivocal and seem sound, yet their effects on the papers themselves are not apparent. Despite producing a steady stream of background research in the area of testing, it has been many years since any bold, substantive changes were made to the FCE. So that such valuable work isn’t looked back on as a missed opportunity, a more reflexive relationship between theory and practice would be advisable, particularly in relation to the fairness and accessibility of test content.

Better articulate the standards set

Too often are Cambridge’s claims regarding fairness expressed as fact and not as aspiration. Much as with the total elimination of cultural bias, 100% test fairness in a global setting can only ever be a commendable target, not an achievable goal in and of itself. It would be best to ensure that these bolder claims are altered to reflect the reality both of the target in mind and the current deficiencies of the tests that aspire to them.
Bibliography


Decock, L. and Horsten, L. (2000) Quine; Naturalized Epistemology, Perceptual Knowledge and Ontology. Amsterdam: Rodopi B. V.


Patrick, F. (2011) Handbook of Research on Improving Learning and Motivation through Educational Games: Multidisciplinary Approaches (pp.1-1462). Hershey, PA: IGI Global


Appendix

Appendix A: Information Sheet ................................................................. 63
Appendix B: Personal Questions Transcript .............................................. 64
Appendix C: Solo Eyes-Closed Transcript ............................................... 65
Appendix D: Personal Picture Description Transcript ............................ 65
Appendix E: 'Museums' Item Transcript ............................................... 66
Appendix F: 'School Activities' Transcript ............................................. 67
Appendix G: Picture Portfolio ................................................................. 68
Appendix H: Cambridge Marking Rubrics .............................................. 72
Appendix I: Ethical Consent Forms .... [Omitted due to British Council rules on branding] .......... 73
Appendix A: Information Sheet

Student Participant Information Sheet

Dear All,

My name is [Omitted] and I am finishing my masters degree at the University of [Omitted]. To complete my studies I am carrying out a research project which will take about three months to complete. You have kindly agreed to consider helping me with my research but it is important that a few things are made clear to you now so that you can make an informed decision. Your indication that you might participate in this project does not mean that you have to help in any way. You can choose not to and, if this happens, you will receive no further contact. However, if you are still able to help, please take the time to read the below carefully.

What is this research about?

This research is concerned with the fairness of an international Cambridge exam called the FCE (First Certificate in English). The purpose is to find areas that may unfairly exclude learners from non-western cultures and suggest improvements. These improvements will be based on making the exam more personal, giving you the opportunity to express your own experience instead of trying to relate to those of others. It hopes to provide some evidence that there are fairer ways of testing that deserve more attention and research.

What will you be asked to do?

You will be asked to do an original FCE speaking test that will take no more than 20 minutes. If you allow it, the test will be recorded on a digital device. Afterwards there will be a short interview about the test. This interview is in no way part of the test itself. It is only interested in your own thoughts and feelings about the test. If you are still happy to continue with the research, you will take another 25 minute test a week later. This test will have different instructions and images. It will also involve some short periods of silence and you will be encouraged to close your eyes if you are comfortable doing this.

How will your information be handled?

All the data I receive from you will be treated confidentially. The only personal details that you will be required to give will be your name and signature on the consent form. At no point will you will not be referred to by name in research itself. The relevant data from any recording you allow me to make will transcribed and then wiped from the original device. The transcriptions will be anonymised and held on a secure University of [Omitted] server. Only myself and the university will be able to access this information. Any physical data collected in paper form during the testing or interview stage will be destroyed. You are welcome to request any of the information you have given me at any point during the project.

How do you withdraw from or complain about this project?

You are not obliged to participate in this research and can withdraw from it at any time. It is important that you feel comfortable with your role in this research and free to leave it if you want to. If you choose to withdraw, you will not be asked for an explanation and will receive no further contact. Any information you have given up to that point will be used only with your permission, or destroyed if you prefer.

If there are any problems with your participation and you feel uncomfortable with the research’s content, then please let me know or contact my course director using the details below (Omitted). If the problem continues or you would prefer to take this to someone else, you can contact the Research Governance Officer on [Omitted] or email at [Omitted] The University of [Omitted] takes such complaints very seriously. When making any complaint, please mention my name and the details of the problem you are having so it can be responded to swiftly.

If you have any questions in the meantime you can contact me using the details at the bottom of this page or my course director, here:

[Omitted]

Thank you for taking the time to read the above and for considering helping me with my research.

[Omitted]
Appendix B: 'Personal Questions' Transcript

Examiner: What do you usually do in the evenings?
Candidate A: Ah.. I go to my friend's. Go to my friends. Go out.. sometimes to work [walk?]. I... or we go to Shisha.
Examiner: Ah, okay.
Candidate A: Yeah, it's typical.. kind of.. for guys to do these things.
Examiner: That's great. And what do you do in the evening? [directed at Candidate B]
Candidate B: Its depends like, at the weekend I'm all the time out. From the morning to the evening. For normal week? - at the gym. I spend my time.
Examiner: Okay, that's great. Umm.. Do you ever go to the cinema?
Candidate B: Yeah, like every weekend.
Examiner: In this country or in Kuwait?
Candidates A and B: Everywhere.
Examiner: Everywhere?
Candidate A: ... because they don't like.. you know.. like..
Candidate B: ... to mix girls with men.
Candidate A: Yeah... Yeah...
Candidate B: Like the Muslim, together.
Examiner: Okay. In the cinemas?
Candidate A: Yeah, [unintelligible] they don't... because it will lead to mix between males and females.
Examiner: Could you tell me about a film you've seen recently, that you enjoyed?
Candidate B: [unintelligible] I don't remember the name but.. went to cinema.. you and me and the guys..
Candidate A: Yeah, what was the movie? Errr.. 'the Rock', you know 'the Rock'? 'The Rock', 'the rock' - the actress was 'the Rock'.
Candidate B: I don't remember the name. It's a new movie for him.
Candidate A: Yeah..
Examiner: So it was like..
Candidate A: He was in it, he was in it..
Examiner: Like an action film?
Candidate A and Candidate B: Yeah, an action film.
Examiner: Good.
Candidate B: It was really great.
Examiner: Umm.. What do you like doing on holiday?
Candidate A: Actually we like to do like new things but we don't have enough information in Liverpool like what the best we should visit or.. so we do normal things. Just go out with friends.. walk.. like this. Yeah 'cos... it's very difficult to make English friends because they know better than us here.. what they should do.
Examiner: Oh, okay. I understand. And you?
Candidate B: Uh, I like to spend my holiday with my family. 'Cos.. on normal days everyone has a work so we don't meet each other all the time, so on holiday everyone is free. So we can stay all the time together.
Examiner: Okay, that's great.
Appendix C: 'Solo Eyes- Closed Alternative Item' Transcript

Examiner: [period of prolonged silence] Can you? Anywhere.. anywhere you feel most comfortable.

Candidate B: My father's farm, I've got like a farm in Kuwait.. mostly, like.. every Kuwaiti has a farm. Once I go at night.. quite late.. no one there.. and go to touch my horse. He's called 'Zeirun'. Stay with him. Sometime watch him.. Clean.. Ride the horse for one hour.. two hour.. No one there. Alone... Don't think about anything.. just riding the horse. Sometimes I talk to him. It's my best time.

Examiner: Yeah, tell me about the horse.. I want to 'see' the horse again.

Candidate B: I treat the horse by different ways. I treat the horse like a 'guest'. It's like a guest, because the horse I think like someone you all the time care about. Like, one month, I didn't come to my farm to see my horse. He's all the time like - the guys that work there - all the time he's just 'shouting', 'shouting' and when I came, he calmed down. Just what I'm like. No one can ride him. No one can ride him. No one..

Examiner: Except you?

Candidate B: Except me.

Examiner: What can you.. what can you hear at night?

Candidate B: Just 'air'

Examiner: Just air, nothing?

Candidate B: ..and the sound from the horse. It's really amazing.. amazing!

Appendix D: 'Personal Picture Description' Transcript

[period of silence]

Candidate A: Ah.. I think it's the best time, like errr.. or the best favourite place where we eating like errr.. in my home... We sit, me and my brothers... my sisters and errr.. my brothers wives.. their children.. my father, mother and we eat and errr.. Before we start eating we have.. you know, like 'argument' and laugh together and errr and have a nice time and when, actually during the eat.. we also eat and errr speak. After that we don't move like, we keep speaking and laughing and together and telling each other what happened this day and I think this is the best moment like in my life.

Examiner: What can you.. what would you eat?

Candidate B: It's rice with chicken, rice with lamb.

Candidate A: Yeah, it's with yellow rice and chicken.

Candidate B: It's from Yemen. Every.. all the rice, all the chicken, of course, is coming from Yemen..

Examiner: Yes?

Candidate B: ...it's become a tradition for us as well.

Examiner: Okay

Candidate B: ... because we're used to it.

Candidate A: Yeah, and we like errr.. it depends on the food because every time we go there is bread, rice - white, yellow and errr.. every day.

Candidate B: Every day for lunch, we go have rice.

Candidate A: And this I think.. we call it 'Durahiya'. It's a place where man sit.

Candidate B: Yeah, I love it!

Examiner: That's good

Candidate B: Fantastic!
Candidate A: Sometimes fish... sometimes fish, sometimes hamburger [unintelligible]. The best thing... we sit together and we speak and that time not allow us to use our phones... you know, and I think that's the best time or the best... favourite place.

Candidate B: It's like a private place with people you love... really love.

Examiner: Of course.

Candidate B: ...and there's like, a relationship with everyone. Like.. 'love'. You don't think like you 'have a job' or anything.. just.. but we don't care about outside because we sit inside the living room and there is air conditioning everywhere.

Candidate B: It's a traditional place. I think you hear about them in Saudi Arabia. It's a traditional place for the man.. the meeen, the men.. and they stay there and they try to talk.

Candidate A: This my brother.. this like is very common place my home. All the time spend[?] most my time there. Can sit there, eat there, sleep there. Every... All the time. During I sit at home I sit there. I think it's the most comfortable place for me and my brothers also.

Examiner: So what are they doing here now? What *might* they be doing? Are they just talking?

Candidate A: I think that now they are fasting. Just sit and read and, and then prayer. Some stuff like that, but in this space it's like all guys they sit there and speak and they can't eat right now (Ramadan). They watch TV, they speak. If they want to sleep also, they sleep there. They just bring cover and pillow. I think that very common place. They all the time sit there and sometimes, like my friends, they sleep in my home like in this space we'll be together - all the time together. Because we don't have responsibility as we will have like in future.. married, children.. and we enjoy our time there. Yeah, everything there - everything! If we have problem we sit there, it's the best place.

Examiner: Okay, perfect

Appendix E: 'Museums' Transcript

Examiner: [instructions given]... [directed at Candidate B] So... can we start here?

Candidate B: Errr..... It's difficult to tell you because I have never never visit a museum.

Examiner: Ah.

Candidate B: I'm not 'terested in that

Examiner: That's okay. No problem at all.

Candidate B: Depends on the person, you know. It's like.. it's interesting to some people.

Examiner: Maybe it's art or.. I don't know.

Candidate B: Okay.. You do not like..

Candidate B: Ummm, I'm not 'trested in that. I'm 'trested in gyms, sports.. these kinds activities.

Examiner: Okay.. okay.. So why *might* someone like those museums?

Candidate B: Ummm, maybe someone 'trested in dinosaurs? .. to see the [names?] and the [bones]?

Examiner: Ummm.. or he do research - like you?

Candidate B: Maybe.

Candidate B: Kind of..

Examiner: Maybe in History?

Candidate B: Okay.. yeah.

Examiner: [pointing to the gallery picture]: Maybe he has a picture of like.. which explains the famous peoples of like actors, presidents..

Candidate B: Maybe... Like photographs, or art?

Candidate B: Ummm.. [shakes his head]

Examiner: No, that's fine.
Appendix F: 'School Activities' Transcript

Examiner: Now I'm going to show you two photos to discuss together, if possible. There are these photos. Actually, there are seven photos here.

Candidate B: Interesting

Examiner: Just take a moment to look at those photos... Here are some different activities that students often do during the school day.

Examiner: [Candidates take some time to look at the photographs]

So, why might it be important... to do these activities at school?

Candidate A: Errr, I think because, to discover like, each person has abil... err.. skills in a part of life, many activities and to discover this err, they make student try all kind of as possible. Here, computers, sport.. I think this like reading or sing poem or something like that.. cooking and these traditional things, and, to discover like [unclear - everything?] Or to give them a bit also like skills, like they might need it in future or something like that.

Examiner: [directed at Candidate B]: Okay, do you agree?

Candidate B: I agree with my friend. I want to add something. Normally at school we start from the morning and afternoon like six or seven hours.. So we can't say like six or seven hours all the students want to study. They should have like some free classes.. some activities to do. Like some people they love video games, they can't have.. classrooms, computer rooms can't spend that.. Also, there is some people.. whatever, girls or female or male or female - they love to cook. Like when I was in high school they have.. there is a free class, you have like music to record or you wanna learn something there is like two or three teachers, they help us.. also there is free like, between classes there is a free time, like a break.

Examiner: Maybe they are singing. Yes.

Examines: Which two activities do you think are not so important?

Candidate A: Not important?.. I think singing and dancing.

Candidate B: Actually, all of them I think is important. I can't say anything is so important, but that one I.. I don't have any idea about, what's that actually? [shrugs shoulders] So I can't say if it's important or not.

Examiner: I think maybe it's acting or theatre.. or 'drama'.

Candidate B: It might be, but in my culture, in my countries - we never ever have it.

Examiner: ..and what was your favourite subject at school? Always sport?

Candidate B: Always

Candidate A: I like.. well like, errr 'Math', especially the teacher like. The teacher makes you like the subject. You will be good and you will love the subject.
Appendix G: Picture Portfolio
(NB: Some Pictures Feature in Multiple Publications)

Sheet 1

Why is the music important to the different groups of people?

What might the people find interesting about visiting these different museums?

Sheet 2

How do the people feel in these situations?

What are the people enjoying about spending time outside in these situations?

Cambridge FCE Teacher's Handbook 2008
(Cambridge: 2008)

Cambridge FCE for Schools Paper 2010
(Cambridge: 2010)

Cambridge FCE Workbook 2008
(Thomas and Thomas: 2008)

FCE Practice Tests Plus
(Kenny and Luque-Mortimer: 2008)

Cambridge FCE for Schools Paper 2015
(Cambridge: 2015)

Cambridge FCE Workbook 2008
(Thomas and Thomas: 2008)
Sheet 3


Sheet 4

Cambridge FCE 2015 Sample Paper (Cambridge English: 2015c)
FCE Practice Tests Plus (Kenny and Luque-Mortimer: 2008)
Cambridge FCE 2015 Sample Paper (Cambridge English: 2015c)
## Appendix H: Cambridge Speaking Assessment Rubrics

<table>
<thead>
<tr>
<th>B2</th>
<th>Grammar and Vocabulary</th>
<th>Discourse Management</th>
<th>Pronunciation</th>
<th>Interactive Communication</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Shows a good degree of control of a range of simple and some complex grammatical forms. Uses a range of appropriate vocabulary to give and exchange views on a wide range of familiar topics.</td>
<td>Produces extended stretches of language with very little hesitation. Contributions are relevant and there is a clear organisation of ideas. Uses a range of cohesive devices and discourse markers.</td>
<td>Is intelligible. Intonation is appropriate. Sentence and word stress is accurately placed. Individual sounds are articulated clearly.</td>
<td>Initiates and responds appropriately, linking contributions to those of other speakers. Maintains and develops the interaction and negotiates towards an outcome.</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>Performance shares features of Bands 3 and 5.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Shows a good degree of control of simple grammatical forms, and attempts some complex grammatical forms. Uses a range of appropriate vocabulary to give and exchange views on a range of familiar topics.</td>
<td>Produces extended stretches of language despite some hesitation. Contributions are relevant and there is very little repetition. Uses a range of cohesive devices.</td>
<td>Is intelligible. Intonation is generally appropriate. Sentence and word stress is generally accurately placed. Individual sounds are generally articulated clearly.</td>
<td>Initiates and responds appropriately. Maintains and develops the interaction and negotiates towards an outcome with very little support.</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>Performance shares features of Bands 1 and 3.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Shows a good degree of control of simple grammatical forms. Uses a range of appropriate vocabulary when talking about everyday situations.</td>
<td>Produces responses which are extended beyond short phrases, despite hesitation. Contributions are mostly relevant, despite some repetition. Uses basic cohesive devices.</td>
<td>Is mostly intelligible, and has some control of phonological features at both utterance and word levels.</td>
<td>Initiates and responds appropriately. Keeps the interaction going with very little prompting and support.</td>
</tr>
<tr>
<td>0</td>
<td></td>
<td>Performance below Band 1.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

## Global Achievement

<table>
<thead>
<tr>
<th>B2</th>
<th>Global Achievement</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Handles communication on a range of familiar topics, with very little hesitation. Uses accurate and appropriate linguistic resources to express ideas and produce extended discourse that is generally coherent.</td>
</tr>
<tr>
<td>4</td>
<td>Performance shares features of Bands 3 and 5.</td>
</tr>
<tr>
<td>3</td>
<td>Handles communication on familiar topics, despite some hesitation. Organises extended discourse but occasionally produces utterances that lack coherence, and some inaccuracies and inappropriate usage occur.</td>
</tr>
<tr>
<td>2</td>
<td>Performance shares features of Bands 1 and 3.</td>
</tr>
<tr>
<td>1</td>
<td>Handles communication in everyday situations, despite hesitation. Constructs longer utterances but is not able to use complex language except in well-rehearsed utterances.</td>
</tr>
<tr>
<td>0</td>
<td>Performance below Band 1.</td>
</tr>
</tbody>
</table>

(Cambridge English: 2015c)