Ontology-Based E-Assessment for Accounting

DRAFT Final Project Report, 23 October 2012

Kate Litherland, Liverpool John Moores University, UK
Patrick Carmichael, University of Stirling, UK

Contents

1. The Project
   1.1 Introduction 2
   1.2 Project Organisation and Affiliations 2
   1.3 Project Aims, Objectives and Deliverables 3
   1.4 Project Progress and Outcomes 3

2. Ontology-Based E-Assessment with OeLe
   2.1 Overview of OeLe's functions and related terminology 6
   2.2 Implementation of OeLe 9

3. Findings, Discussion and Recommendations
   3.1 Project Findings 10
   3.2 Discussion 13
   3.3 Future Directions 15
1. The Project

1.1 Introduction

The ‘Ontology-Based E-Assessment for Accounting’ project has implemented and explored the potential of a novel ontology-based e-assessment system that draws on the potential of emerging semantic technologies to produce an online assessment environment that is capable of marking students’ free text answers to ‘conceptual’ rather than factual, multiple-choice questions. The system used, OeLe, does this by matching student response with a ‘concept map’ or ‘ontology’ of domain knowledge expressed and vetted by subject specialists.

OeLe supports automated marking, and can also be integrated with systems that allow for feedback to individual students about their strengths and weaknesses and recommend resources to support further learning and revision, as well as providing tutors with information on individuals and whole cohorts, thus providing both a formative as well as summative function. This final report details the first implementation and evaluation of an OeLe based e-assessment system in the context of an undergraduate course in Financial Accounting, in which the automated marking aspects of the system were evaluated.

The report describes the potential affordances and demands of implementing ontology-based assessment in Accounting. It considers the various ways in which ontology-based e-assessment might be used to support human markers and considers the implications of using the system in these ways for students, teachers, and examining bodies. The report concludes with some suggestions of future directions of work and research if ontology-based e-assessment approaches are to be more widely implemented in Accounting education.

1.2 Project Organisation and Affiliations

The project was based at Liverpool John Moores University, UK, and was directed by Professor Patrick Carmichael (now at the University of Stirling, UK) and Dr. Kate Litherland. Other investigators, based at the University of Murcia, Spain were Professor Maria Paz Prendes, Dr. Jesusaldo Tomas Fernandes-Breis and Dr. Maria del Mar Sanchez. Other members of the project team were Agustina Martinez-Garcia and Rob Crichton.

The project was funded by ACCA (Association of Chartered Certified Accountants) and the International Association for Accounting Education and Research (IAAER) under a programme of research to support the work of IFAC’s International Accounting Education Standards Board (IAESB).
The project commenced in April 2011 and reported interim findings in June 2011 and October 2011, and also met ACCA staff in London in September 2011. This report updates those presented at these previous meetings and also presents findings and recommendations.

1.3 Project Aims, Objectives and Deliverables

The project's main research question was ‘What is the potential of ontology-based semantic technologies for the summative and formative e-assessment of undergraduate learning in accounting?’

In order to address this question, we set out the following research objectives:

- **RO1**: To identify suitable areas for ontology-based e-assessment of undergraduate accounting, and to express these as a formal ontology.
- **RO2**: To identify existing resources to support student learning of accounting and to construct these into a semantically marked up collection to integrate with the ontology-based e-assessment tool used in the project, OeLe.
- **RO3**: To develop a set of assessment activities involving extended student writing which relate to threshold concepts, the project’s formal ontology and existing pedagogical support materials.
- **RO4**: To deploy an instance of the OeLe e-assessment system, and to carry out a technical and pedagogical evaluation of the system.
- **RO5**: To document and share project materials and processes by which others can develop their own instances of OeLe, as well and disseminating the project's findings and evaluations.

The project also proposed to disseminate project documentation and findings through both professional and academic networks, and to produce articles for relevant journals.

1.4 Project Progress and Outcomes

By September-October 2011, we had largely met the first three of our research objectives, although we had encountered some difficulties with these, as we described in previous reports, and needed to make strategic decisions about how to proceed - leading to our concentrating our efforts on developing and evaluating the core automatic marking functions of the system.

Now, in October 2012, we can additionally report on the results of our work towards the latter two research objectives (RO4 and RO5). Reviewing each objective in turn provides a good overview of project activities ahead of discussion of findings.
RO1: Identify suitable areas for ontology-based e-assessment of undergraduate accounting, and to express these as a formal ontology.

We identified a suitable area of undergraduate study, a Year 2 module in Financial Accounting, one of the first in which students encounter the conceptual basis of the subject. Discussions with teaching staff suggested that students found this aspect of the course particularly challenging.

The conceptual basis of this course was expressed in a formal ontology, created using Protégé, a free tool available at http://protege.stanford.edu. Protégé is a highly sophisticated tool, but is very challenging for non-specialists to use even with extensive support - this informs one of our recommendations, namely that any future work on ontology-based e-Assessment will require ontology authoring tools that are easier to use for teachers and course designers without the support of a specialist in knowledge management and representation.

RO2: Identify existing resources to support student learning of accounting and to construct these into a semantically marked up collection to integrate with OeLe.

Having identified the setting for the trial, we investigated existing support materials which might be integrated with OeLe as online materials to be presented to students encountering difficulties. However, what we discovered was a strong reliance on textbooks and paper materials created by the tutor (e.g. worked problems), coupled with reluctance on their part to engage with other online materials - developing new course content as well as new assessment practices was seen as too much change, too fast. As this aspect of OeLe is not a technically challenging one to implement, we decided to focus efforts instead on the more demanding work relating to fully automated marking of exams.

RO3: Develop a set of assessment activities involving extended student writing which relate to threshold concepts, the project’s formal ontology and existing pedagogical support materials.

Concerns about changes to the experience of students on the module this year in comparison to others, coupled with the findings outlined above, lead us to adopt pre-existing questions for use in trials rather than introducing new assessment activities. We developed ontologies and OeLe examinations to replicate existing short-answer tests offered in the Financial Accounting module between 2006 and 2010 and offered these to students in 2011 as formative assessments prior to their own test.
RO4: Deploy an instance of the OeLe e-assessment system, and carry out a technical and pedagogical evaluation of the system.

At the beginning of the project, the team from Murcia travelled to Liverpool John Moores University to familiarise the UK team with the system. We subsequently translated the system from Spanish into English and installed the system at LJMU. This was then used as the basis of the evaluation that is reported in the ‘Findings’ section of this report.

RO5: Document and share project materials and processes by which others can develop their own instances of OeLe, as well as disseminating the project’s findings and evaluations.

As part of the ‘live’ trials with students and teachers described above, we created documentation to support students, markers, and administrators of the system (the version using manual annotation and automatic marking).

These guides are now available to download on our website, which also carries a short description of the project and relevant links (http://ensembleljmu.wordpress.com). The issues with Protégé described above meant that we did not document the process of ontology and exam creation, as this part of the process still requires specialist support, and we are continuing to explore alternatives to the use of specialised software or experts. Our findings from the trials of OeLe have been reported in an article for *Journal of Accounting Education*, which has been accepted for publication (subject to minor revisions) in December this year.
2. Ontology-Based E-Assessment with OeLe

2.1 Overview of OeLe’s functions and related terminology

OeLe has at its core is a model of the domain knowledge or course ontology to be assessed. This is associated with, and may be designed alongside, an examination, comprising a set of questions and model answers that accompany them, and which teachers need to develop, along with marking and weighting criteria. Each question is assigned a number of marks, as in any examination, but additionally, the relative values of the different concepts that appear in the model answers are also assigned ‘weights’. This allows teachers to assert that for a particular question, it is more important that students recognise the salience of one concept than another. The OeLE system can be represented visually as shown in Figure 1.

![Figure 1: Overview of the OeLE system](image)

Standards and course content inform not only the design of examinations (as would normally be the case) but also the development of model answers and the course ontology - a ‘map’ of the concepts that the examination is designed to assess and the weightings to be attached to concepts. At this stage, any automated assessment would be based on matching student answers with the exact terms that appear in the course ontology and scores based on the values of the questions and the weights of the concepts would be calculated.

As all students are unlikely always to express their answers in the precise language that appears in the ontology, a range of acceptable alternative linguistic expressions may be defined.
that are mapped to the concepts in the ontology. The initial source of these is the model answer, but OeLe can also be ‘trained’: as student answers are assessed and annotated, markers can highlight additional acceptable linguistic expressions and associate them with concepts in the ontology, so that subsequent student answers can be assigned marks even though their responses may not exactly match the model answers. One of the key questions for the project, and for OeLe development more generally, is the amount of training required to allow reliable automated marking without teacher intervention.

Students submit their answers through a web interface which resembles other online examination and survey tools; examinations can include both closed (multiple choice) and open (text response) answers and can be opened for a set time period during which students may either make a single attempt to answer the questions, or return to revise their answers at any time during the examination period. This latter option offers the possibility of students being presented with ‘open book’ style questions on which they work over a period of time until they are satisfied with their answers, or in more reflective assessments in which they progressively elaborate their response.

The system can therefore be used in two modes: in the first, teachers carry out annotation through an interface which allows them to read student answers, highlight excerpts and associate these with the relevant concepts. The OeLe system then calculates student scores according to values and weightings, but this process also enables the training of the system so that it can be used to process the text of student answers, identify exact matches or acceptable alternatives, and then calculate scores on the basis of the values and weightings as before. For the remainder of this report we will describe the first mode as automatic marking (with manual annotation) and the latter as automatic annotation, the latter also calculating scores.

Once marking is complete, students are then able to receive feedback derived from the same ontology that underpins the annotation and marking processes. This includes their mark, the model answer to compare with their own, and a summary of the concepts for which they received credit and a list of concepts which they could have drawn on to receive higher mark. This list of concepts may then be linked to suggestions of useful resources or revision activities that might help to develop understanding (see Figure 2).

Teachers’ general feedback to a student cohort, too, can be couched in terms of understanding and application of concepts rather than success in answering specific questions. Teachers also receive feedback about their success in conveying the conceptual basis of their course content, but also how well the assessment exercises they have set are indeed testing conceptual understanding (Figure 3).
Figure 2: Student Feedback on Concepts Relevant to a Question

Figure 3: Teacher Report on Best And Least Understood Concepts Across a Student Cohort
2.2 Implementation of OeLe

As outlined above, our implementation and subsequent work focused on the automatic marking and annotation functions of the OeLe system (the portion of Figure 1 shaded in grey). While we were able to generate teacher reports as shown in Figure 3, we did not fully implement this aspect of the system, which has the potential to generate a range of reports, charts and other representations of individual and cohort performance, and focussed our attention on the questions of accuracy of annotation and marking.

Following a series of trials with small sets of test answers, between October 2011 and June 2012, we carried out the following tests of the system based on sets of examinations completed by students in the Second Year Undergraduate course in Financial Accounting.

- A trial using 30 marked scripts, to ascertain how best to configure the system for Accounting, implement the ontology, and test the automatic marking of manually annotated scripts
- Live trials of four sets of OeLe tests with students, as formative self-assessment, using manual annotation and automatic marking.
- A trial of the fully automated system (automatic annotation + automatic marking) with the same 30 marked scripts used in the pilot.
- A trial using all 103 marked scripts to assess the potential of the fully-automated system to annotate and assign marks; to ascertain how much ‘training’ the system needed to do this consistently; its ability to deal with different types of questions and responses; and its accuracy and predictability in comparison to a human marker

We will report on the first and fourth of these in detail here.

As the project proceeded, we found that some of the use-cases that emerged from our work were not fully supported by the OeLe system and that other aspects of the OeLe system (which was originally developed to assess student learning in education) did not map well to the kinds of questions, answers and marking strategies we needed to use it effectively in Financial Accounting. As a result members of the project team at LJMU and Murcia spent more time working to develop and adapt the system than we had originally anticipated.
3. Findings, Discussion and Recommendations

3.1 Project Findings

The first trial was based on a sample set of 30 papers, and was designed to ascertain how best to configure the system for Accounting, implement the ontology and test the automatic marking of manually annotated scripts. This involved comparison of the manual marks (‘pencil and paper’ style) with those achieved by a marker reading and annotating each of the scripts using the terms in the ontology, with OeLe then calculating the marks to be awarded.

Table 1: Comparison of manual and auto-annotated/marked scores across Q1-6 on 30 ‘sample’ papers

<table>
<thead>
<tr>
<th>Question (max mark)</th>
<th>Manual marker</th>
<th>Manual annotation, automarked</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>1 (3)</td>
<td>1.97 (1.10)</td>
<td>1.38 (1.04)</td>
</tr>
<tr>
<td>2 (2)</td>
<td>1.23 (0.97)</td>
<td>1.06 (0.87)</td>
</tr>
<tr>
<td>3 (3)</td>
<td>0.73 (0.98)</td>
<td>0.80 (0.75)</td>
</tr>
<tr>
<td>4 (2)</td>
<td>0.57 (0.73)</td>
<td>0.60 (0.67)</td>
</tr>
<tr>
<td>5 (5)</td>
<td>2.10 (1.40)</td>
<td>1.51 (0.96)</td>
</tr>
<tr>
<td>6 (8)</td>
<td>4.00 (2.63)</td>
<td>4.53 (2.52)</td>
</tr>
<tr>
<td>Totals</td>
<td>10.6 (4.80)</td>
<td>9.88 (4.09)</td>
</tr>
</tbody>
</table>

Using the ontology to guide annotation led to a more focused marking process than the wholly manual marking, as it compelled the marker to highlight text and then assert relationships with concepts from the ontology - in effect, causing them to justify the award of marks rather than placing an indicative ‘tick’ on the script.

The combination of scores (per question) and weightings (per concept) also meant that when marks were calculated, a range of fractional marks were achieved rather than manual marks of integer values (with occasional ‘half marks’). The importance and influence of the model answers became evident in this respect: in Question 2, 17 of the 30 students achieved the maximum 2 marks when marked manually. However, the model answer included a small detail which only 1 student included in their response (thus achieving the maximum 2 marks when manually annotated and auto-marked) while the other 16 students scored 1.78 when auto-marked. Rounding would have led to their resulting reported mark being the same, but this
highlighted the fact that, in questions where very specific responses were required, the ontology-based system had the potential to discriminate in fine detail between answers.

Inconsistency in manual marking of near-identical answers was highlighted when these were compared with their manually-annotated and automatically marked counterparts. For example, students who wrote virtually identical answers to question 5, but received manual marks of 2 and 3 were all awarded a consistent 1.59 when OeLe was used, for example.

Again, by being asked to explicitly indicate which part of the answer related to the concept being credited, markers were compelled to focus on what the student answer actually means, rather than being swayed by style or expression. Markers were at once discouraged from giving marks to concepts which are vaguely expressed, or merely implied, and at the same time, encouraged to recognize and reward detail where it was present.

This highlights the importance of teachers and examiners working to create a sufficiently detailed and accurate conceptual structure at the beginning of the examination process, capturing the required detail and reflecting the various elements which might be present, and independently credited, in student answers.

In the fourth trial we drew on the first, second and third to carry out and evaluate a complete ‘training’ process. In Trial 1, which was concerned with manual annotation, the human marker reads the student answer on-screen, identifies the ideas present and associates these with the relevant parts of the ontology: the answer needs to be precise enough for a specific part of it to be recognizable as the expression of a specific concept, but it does not need to be couched in exactly the same terms as the marker can recognize acceptable synonyms. For OeLe to do the same, it first needed to be ‘trained’ in recognizing these alternative expressions.

‘Training’ consists of annotating some of the students’ answers manually: this enables the system to append them to the concepts already in the ontology. Here, the distinction between a formal, expert ontology, which might contain exact synonyms, and the situated ‘working ontology’ of OeLe, where ‘acceptable answers’ are allowed by teachers in recognition that students might lack the full technical vocabulary of the subject, needs to be kept in mind.

Samples of a student cohort set of examinations (n=103) were manually annotated and then the OeLe system was used to process, auto-annotate and mark the whole cohort. We were concerned to explore how many student examinations needed to be manually annotated in order for the system to annotate and mark the examination without human intervention, and as such we ran ‘training sessions’ with samples of 10, 20 and 30 examinations. The results of this trial are summarised in Table 2.
Table 2: Fully automated annotation and marking (n=103) with OeLe trained on 10, 20 and 30 scripts

<table>
<thead>
<tr>
<th>Q.</th>
<th>Manual marker</th>
<th>Trained on model answer</th>
<th>Trained on model answer plus 10</th>
<th>Trained on model answer plus 20</th>
<th>Trained on model answer plus 30</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.22 (1.22)</td>
<td>0.45 (0.62)</td>
<td>0.78 (0.91)</td>
<td>1.12 (1.06)</td>
<td>1.13 (1.06)</td>
</tr>
<tr>
<td>2</td>
<td>0.80 (0.96)</td>
<td>1.04 (0.79)</td>
<td>1.04 (0.79)</td>
<td>1.04 (0.79)</td>
<td>1.04 (0.79)</td>
</tr>
<tr>
<td>3</td>
<td>0.72 (0.90)</td>
<td>0.91 (0.91)</td>
<td>0.95 (0.89)</td>
<td>0.97 (0.90)</td>
<td>0.99 (0.90)</td>
</tr>
<tr>
<td>4</td>
<td>0.49 (0.62)</td>
<td>0.31 (0.50)</td>
<td>0.35 (0.52)</td>
<td>0.39 (0.56)</td>
<td>0.44 (0.62)</td>
</tr>
<tr>
<td>5</td>
<td>1.79 (1.42)</td>
<td>0.97 (1.10)</td>
<td>1.00 (1.10)</td>
<td>1.09 (1.12)</td>
<td>1.10 (1.12)</td>
</tr>
<tr>
<td>6</td>
<td>2.96 (2.50)</td>
<td>3.07 (2.81)</td>
<td>3.48 (2.92)</td>
<td>3.56 (2.96)</td>
<td>3.62 (2.99)</td>
</tr>
<tr>
<td></td>
<td>7.97 (5.39)</td>
<td>6.75 (4.66)</td>
<td>7.60 (4.98)</td>
<td>8.17 (5.27)</td>
<td>8.31 (5.31)</td>
</tr>
</tbody>
</table>

The scores achieved by automatic annotation against the ontology and model answer (Column 3) are much lower than those awarded by the human marker in the original examination (Column 2). This represents OeLe’s annotation and marking of student answers based only on a single model answer and the associated ontology. There is no room here for ‘discretion’ - no acceptable answers or ‘words to that effect’: and scores are lower on longer questions. Questions 2 and 3 were short-answer questions in which the appearance of specific terminology was essential for marks to be awarded, and here the impact of training was significantly less.

However, as columns 4-6 show, the patterns of scores on other questions became increasingly close to those of the original human marker as the number of training scripts increased, with 20 scripts apparently enough to ensure a good agreement both in terms of the spread of marks shown here and a generally good correlation of original examination scores and full automatic annotation and marking and a tendency towards ‘exhaustion of possibilities’ between n=20 and n=30.

We were interested to explore these findings in more depth: for example, discovering whether high-scoring students continued to score highly when their work was auto-annotated and auto-marked. A Pearson R correlation of 0.84 was achieved between the scores of all 103 scripts awarded by manual (human) marking and the full auto-annotation and automatic marking where the training sample was 30. Correlations of scores on specific questions varied from
0.86 (Question 1) and 0.83 (Question 6) to a low of 0.38 (Question 4), which can be attributed to generally low scores on this question.

When we reported these results to participants, it proved useful to express these findings in terms in which teachers and students couched many of their questions and concerns about the role of e-assessment:

- Of the 103 students, 42 students would have gained marks (after rounding) had a trained OeLe system been used; 41 would have lost marks and the remainder would have emerged with the same mark as the original marker awarded.
- If we assume a bare passing score is around 40%, then 10 students who would have failed this part of the examination would have been awarded a passing grade by OeLe, while 6 would have dropped below the 40% threshold.
- If we assume that a high pass, distinction or ‘first class’ is 70%, 5 students who did not achieve this would have been awarded this grade by OeLe, while a further 6 would have dropped below the 70% threshold as a result of the automatic marking.
- OeLe performed best with marks at the upper and lower end of the range, clearly identifying students with very low or very high marks. Scores within +/-20% of the pass mark were less consistent, but even so, the first point of divergence from the human marker between overall pass and fail marks was the student in 74th position of 103.

The system’s ability to recognize very poor and very good responses may itself have potential, particularly if used to filter out papers significantly below the pass mark before passing them on to (human) markers for either manual annotation or ‘traditional’ manual marking where time and budgetary constraints are important factors to consider when allocating marking.

Indications from this trial suggest that OeLe could be used with reasonable confidence to screen out the lowest quartile of papers, allowing markers to focus their efforts on scripts where some relevant content has already been recognized (or, for formative assessments, to identify those in need of more significant levels of support).

### 3.2 Discussion

The project set out to answer the question ‘What is the potential of ontology-based semantic technologies for the summative and formative e-assessment of undergraduate learning in accounting?’ and we can conclude that:
• By modeling answers and ‘training’ an ontology-based e-assessment system, it is possible to assess students’ understanding of concepts in Accounting with fair degrees of accuracy and consistency.

• Ontology-based e-assessment has potential for use in supporting markers by structuring their practices and enforcing desirable marking practice, while removing from them the ‘judgement calls’ of what numerical integer marks to award.

• Ontology-based e-assessment can apply fine-grained analysis of the concepts articulated in students’ answers in a way that human markers may not be able to do consistently.

• Ontology-based e-assessment may allow systems or markers to identify students at extremes of the marking spectrum, particularly those in the lower quartile of scores.

• Successful implementation of ontology-based e-assessment demands that teachers and examining bodies make clear the conceptual basis of courses and the assessments that accompany them.

Whilst the overall patterns described above are of improved consistency and ‘agreement’ with the manual marker when using a ‘trained system’, this belies what may be a broader and more strategic question about the validity of an approach which primarily aims to replicate human markers’ strategies, rather than being oriented towards greater levels of consistency. As we have indicated, disparities between OeLe and the human marker were most evident where students expressed partial understanding of key concepts. Where students can state key ideas clearly, OeLe rewards their answers, even if their reasoning is incomplete or poorly expressed. In contrast, the reverse is true of the human marker, who tends to reward students who can express the ‘gist’ of a correct response, but in very general and imprecise terms. These divergent interpretations of ‘understanding’ are difficult to reconcile: whilst OeLe operates on the basis that using the correct terminology in the specified context implies understanding, the human marker’s approach is more subtle, but because of this, potentially also more inconsistent (especially where multiple markers are used). Which of these approaches to ‘understanding’ is most congruent with the types of knowledge which students are required to express in assessments is therefore part of a broader discussion about what constitutes professional knowledge in Accounting, and how the competences required for professional practice may best be examined.

Attempting to accurately reproduce the work of a human marker may not, then, be the most fruitful avenue for development of any e-assessment system. Making the most of a system like OeLe implies a different approach to testing, which includes, but is not limited to, changes in
marking practice. Ontology-based e-assessment systems like OeLe can apply fine-grained analysis of the concepts articulated in students’ answers in a way that human markers may not be able to do consistently, but a different approach to the assessment process may be needed in order to fully exploit this potential. In our trials, ‘complete’ responses only got a few tenths of a point more than merely ‘good’ ones, because the test had been designed with existing practices in mind.

We can identify three potential strategies (or combinations thereof) for the further development and implementation of systems such as OeLe:

- A strategy privileging **consistency**: to support greater levels of consistency by supporting markers, training new markers and allowing monitoring and moderation
- A strategy privileging **reproduction of human marker practice**: in which the aim is to provide students experience and outcomes of assessment as close to that as possible to that provided by a human marker (an interesting direction given recent enthusiasm for MOOC’s - Massive Open Online Courses - where automated assessment is seen as a means of supporting very large student cohorts)
- A strategy privileging **efficiency** in relation to assessment processes, as exemplified by the use of OeLe to ‘screen’ out examinations that are incomplete, incomprehensible or otherwise clearly not of ‘passing’ quality and therefore saving markers’ time and their employers’ money.

### 3.3 Future Directions

Our work was always envisaged as a small-scale and early exploration of emerging technologies but even so has generated useful findings. It has also helped us frame broader questions about e-Assessment more generally - in terms of its technological basis but also the pedagogical and assessment practices into which it might be integrated. These fall into four areas (which in turn might be influenced by the strategic roles for e-Assessment systems identified above).

**More Robust Technological Systems**: OeLe has been a useful and functional ‘proof of concept’ tool for showing the potential of ontology-based e-assessment. However, its legacy of features which were designed for a different assessment regime, its reliance on Protégé as a means of generating and structuring ontologies, and aspects of its technical configuration mean that the actual version of the system used in these trials would need considerable work before wider deployment was advisable.

**Scaling Up**: A more robust e-Assessment framework would enable and support experiments with much larger student cohorts. The question of to what extent the overhead of ‘training’
would diminish (remembering that in our trial, a human ‘annotator’ had to mark 30 papers for the system to mark another 73!) is of particular interest here. It may be that this varies widely between courses, level of challenge and ‘openness’ of questions.

**Better Understanding of Assessment Practices:** Implementing OeLe provided an insight into existing pedagogical and assessment practices, particularly in relation to those used by teachers to prepare students for examination. Given that examinations are already ‘performative’ (that is, their design not only reflects but directs pedagogy more generally) the introduction of e-Assessment systems would need to be accompanied by some careful exploration of its effects on teacher and student behaviour. One aspect of this which impacts directly on systems like OeLe (which use both model answers and ontologies) is what roles different representations of knowledge (case studies, worked examples, answers, ontologies, standards, and the examination questions themselves) play - and, indeed, how these are related. Are ‘model answers’, for example, presented as **complete answers** covering every aspect of a conceptual area? Or are they offered to students as **possible, good answers** which would achieve full marks in the examination? Training teachers to write the kinds of complete model answers that OeLe draws on to mark student answers may not currently be the practice amongst all (or even any) teachers.

**Integration with Other Systems:** We have concentrated - in this project and in this report - on OeLe as a ‘free standing’, summative assessment system offering formative feedback. Further work could useful explore how systems such as OeLe could be used more formatively, as part of ongoing student learning and teacher development. Other integration with virtual learning environments, resource banks and student management systems is also a potential area for future developments.