Technology Enhanced Assessment: Do we have a wolf in sheep’s clothing?

Professor Denise Whitelock
The Open University, Walton Hall,
Milton Keynes MK7 6AA, UK
denise.whitelock@open.ac.uk
Assessment
The e-Assessment and automatic feedback Challenge

- Constructivist Learning – Push
- Institutional reliability and accountability – Pull
Grand Challenge representing analysis of learning that can be readily understood
Providing meaningful automatic feedback

• Can we provide Advice for Action as part of eAssessment?
The LISC solution to language learning: developed by Ali Fowler

- A CALL system designed to enable students to:
  - Independently practise sentence translation
  - Receive immediate (and robust) feedback on all errors
  - Attend immediately to the feedback (before fossilisation can occur)
How is the final mark arrived at in the LISC System?

- The two submissions are *unequally* weighted
  - Best to give more weight to the first attempt
    - since this ensures that students give *careful* consideration to the construction of their first answer
    - but can improve their mark by refining the answer
  - The marks ratio can vary (depending on assessment/feedback type)
    - the more information given in the feedback, the lower the weight the second mark should carry
What about emotional support in the feedback?

- Difficult at times to receive written feedback
- Not just a cognitive response
- How can Bales help?
## Coding the tutor comments

### Bales’ Interaction Process

<table>
<thead>
<tr>
<th>Categories</th>
<th>Specific Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Positive Reactions</strong></td>
<td></td>
</tr>
<tr>
<td>A1</td>
<td>1. Shows solidarity</td>
</tr>
<tr>
<td>A2</td>
<td>2. Shows tension release</td>
</tr>
<tr>
<td>A3</td>
<td>3. Shows agreement</td>
</tr>
<tr>
<td></td>
<td>Jokes, gives help, rewards others</td>
</tr>
<tr>
<td></td>
<td>Laughs, shows satisfaction</td>
</tr>
<tr>
<td></td>
<td>Understands, conurs, complies, passively accepts</td>
</tr>
<tr>
<td><strong>Attempted Answers</strong></td>
<td></td>
</tr>
<tr>
<td>B1</td>
<td>4. Gives suggestion</td>
</tr>
<tr>
<td>B2</td>
<td>5. Gives opinion</td>
</tr>
<tr>
<td>B3</td>
<td>6. Gives information</td>
</tr>
<tr>
<td></td>
<td>Directs, proposes, controls</td>
</tr>
<tr>
<td></td>
<td>Evaluates, analyses, expresses feelings or wishes</td>
</tr>
<tr>
<td></td>
<td>Orients, repeats, clarifies, confirms</td>
</tr>
<tr>
<td><strong>Questions</strong></td>
<td></td>
</tr>
<tr>
<td>C1</td>
<td>7. Asks for information</td>
</tr>
<tr>
<td>C2</td>
<td>8. Asks for opinion</td>
</tr>
<tr>
<td>C3</td>
<td>9. Asks for suggestion</td>
</tr>
<tr>
<td></td>
<td>Requests orientation, repetition, confirmation, clarification</td>
</tr>
<tr>
<td></td>
<td>Requests evaluation, analysis, expression of feeling or wishes</td>
</tr>
<tr>
<td></td>
<td>Requests directions, proposals</td>
</tr>
<tr>
<td><strong>Negative Reactions</strong></td>
<td></td>
</tr>
<tr>
<td>D1</td>
<td>10. Shows disagreement</td>
</tr>
<tr>
<td>D2</td>
<td>11. Shows tension</td>
</tr>
<tr>
<td>D3</td>
<td>12. Shows antagonism</td>
</tr>
<tr>
<td></td>
<td>Passively rejects, resorts to formality, withholds help</td>
</tr>
<tr>
<td></td>
<td>Asks for help, withdraws</td>
</tr>
<tr>
<td></td>
<td>Deflates others, defends or asserts self</td>
</tr>
</tbody>
</table>
Identifying trends: H801

Graph to show conflated Bale’s categories against mean number of incidences in H801 scripts
Welcome to Open Mentor

You're here: OpenMentor >> H804 >> View reports

OpenMentor comment analysis
This page shows details of how each assignment's comments have been analysed.

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Course</th>
<th>Assignments marked</th>
</tr>
</thead>
<tbody>
<tr>
<td>TMA 01</td>
<td>H804</td>
<td>25</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Comment category</th>
<th>Guidelines</th>
<th>Your comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category A: Positive feedback</td>
<td>-</td>
<td>It is very good practice to read through your work...</td>
</tr>
<tr>
<td></td>
<td></td>
<td>... Good idea</td>
</tr>
<tr>
<td>Category B: Teaching points</td>
<td>More than expected</td>
<td>This could have done with a little academic distance eg you...</td>
</tr>
<tr>
<td></td>
<td></td>
<td>You might need to be a little more explicit: technology...</td>
</tr>
<tr>
<td>Category C: Questions</td>
<td>-</td>
<td>Are these your fellow students on H804?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.. Did you get this from a particular report?</td>
</tr>
<tr>
<td>Category D: Negative reactions</td>
<td>More than expected</td>
<td>.. not at all scalable</td>
</tr>
<tr>
<td></td>
<td></td>
<td>You do not need to put the page numbers in here</td>
</tr>
</tbody>
</table>

- Show the full comments
- Show this as a bar chart
- Give me feedback on my use of comments
Welcome to Open Mentor

You're here: OpenMentor >> H804 >> View reports

OpenMentor comment analysis
This page shows details of how each assignment's comments have been analysed.

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Course</th>
<th>Assignments marked</th>
</tr>
</thead>
<tbody>
<tr>
<td>TMA 01</td>
<td>H804</td>
<td>25</td>
</tr>
</tbody>
</table>

Assessment for Sandy Smith: mark assigned 61 (Grade 3)

Bar charts to represent analysis of assignments sent to Open Mentor

Category keys table

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category A</td>
<td>Positive Reactions</td>
</tr>
<tr>
<td>Category B</td>
<td>Teaching Points</td>
</tr>
<tr>
<td>Category C</td>
<td>Questions</td>
</tr>
<tr>
<td>Category D</td>
<td>Negative Reactions</td>
</tr>
</tbody>
</table>

● Show this as a summary table
What can we learn from modelling tutors marking to construct a formative e-assessment tool?

- Learning from pedagogical practice to build an automated tool
- Open Comment project builds on the work of OpenMentor
- Free text entry for History and Philosophy students
- Immediate feedback (in context) to students
- Influenced by ELIZA (Weizenbaum, 1963)
Open Comment addresses the problem of free text entry

- Automated formative assessment tool
- Free text entry for students
- Automated feedback and guidance
- Open questions, divergent assessment
- No marks awarded
- For use by Arts Faculty

DMW AHE Conference June 2017
Evidence from literature re Praise feedback for Open Comment

- Praise for ability per se can hinder learning (Mueller & Dweck, 1998)
- Praise = being clever
- Negative feedback now without ability
- Disempowering and demoralising

DMW AHE Conference June 2017
Mueller & Dweck (1998)

- Raven’s Matrices (IQ)

- First test pupils praise either for effort or ability

- Second test most difficult

- Third test medium difficulty. Score up 1 points for pupils praised for effort. Down 1 point ability
How does feedback effect mindsets?

1. Your intelligence is something very basic about you that you can’t change very much

2. You can learn new things but you can’t really change how intelligent you are

3. No matter how much intelligence you have you can always change it quite a bit

4. You can always substantially change how intelligent you are
Mindsets (Dweck, 2006)

Fixed mindset

- Super sensitive about being wrong
- Always trying to prove themselves

Growth mindset

- Stretch themselves
- Confront obstacles as challenges
- Lack of tension when learning as they know they are novices and can improve
Open Comment Arts

You are logged in as Stuart Watt (Logout)

Moodle > AA000 > Quizzes > Combined Quiz > Review

Info  Results  Preview  Edit

Click here to go back to the course

1

Read the first paragraph in which Joll analyses why the outbreak of the First World War has been studied in such detail. What reasons does he give?

Answer: No idea

Feedback:

Maybe you are a bit confused by the question. It may be helpful to remember you are not being asked about the causes directly, but why the causes have been so extensively studied.
Stages of analysis by computer of students’ free text entry for Open Comment: advice with respect to content (socio-emotional support stylised example)

- **STAGE 1a: DETECT ERRORS** E.g. Incorrect dates, facts. (Incorrect inferences and causality is dealt with below)
  - Instead of concentrating on X, think about Y in order to answer this question. Recognise effort (Dweck) and encourage to have another go
  - You have done well to start answering this question but perhaps you misunderstood it. Instead of thinking about X which did not........ Consider Y
Computer analysis continued

- **STAGE 2a: REVEAL FIRST OMISSION**
  
  Consider the role of Z in your answer. Praise what is correct and point out what is missing. Good but now consider the role X plays in your answer.

- **STAGE 2b: REVEAL SECOND OMISSION**
  
  Consider the role of P in your answer. Praise what is correct and point out what is missing. Yes but also consider P. Would it have produced the same result if P is neglected?
Final stages of analysis

- STAGE 3: REQUEST CLARIFICATION OF KEY POINT 1
- STAGE 4: REQUEST FURTHER ANALYSIS OF KEY POINT 1 (Stages 3 and 4 repeated with all the key points)
- STAGE 5: REQUEST THE INFEERENCE FROM THE ANALYSIS OF KEY POINT 1 IF IT IS MISSING
- STAGE 6: REQUEST THE INFEERENCE FROM THE ANALYSIS OF KEY POINT 1 IF IT IS NOT COMPLETE
- STAGE 7: CHECK THE CAUSALITY
- STAGE 8: REQUEST ALL THE CAUSAL FACTORS ARE WEIGHTED
SAFeSEA: Supportive Automated Feedback for Short Essay Answers

An automated tool supporting online writing and assessment of essays providing accurate targeted feedback

SAFeSEA
Professor Denise Whitelock
Professor John Richardson
Professor Stephen Pulman

DMW AHE Conference June 2017
About SAFeSEA

• No tutor support for drafts of first assignment
• Reduce dropout rate with automatic feedback?
• Effect of summaristion
• What are the beneficial factors?
• Correlate measures of learner activity and essay improvement
• http://www.open.ac.uk/iet/main/research-innovation/research-projects/supportive-automated-feedback-short-essay-answers
The system’s focus is to present summaries of students’ own work in different ways, to encourage them to reflect constructively on what they have written.

In other words Open Essayist tells them from its analysis what are the most important or key points in their essay. They can then think about whether that was what they intended to emphasise in their essay. If not then they can make the appropriate changes.

A very important aspect of the OpenEssayist system is that it will not tell students what to write, or how to rewrite sections of their essay, or even what is correct or incorrect in their essay.
OpenEssayist: How it gives feedback

• Three aspects of the students’ essays are analysed by the system:
  • the **structure** of the essay (which paragraphs constitute the introduction, the conclusion, the discussion sections, etc.),
  • the **key words and key phrases** of their essay (which are the most important words and phrases, the ones that are most representative of the essay's overall meaning)
  • the **key sentences** of their essay (which are whole sentences that are most representative of the essay's overall meaning).
Did I really mean that? Applying automatic summarisation techniques to formative feedback

Authors: Debora Field, Stephen Pulman, Nicolas Van Labeke, Denise Whitelock, John TE Richardson

Abstract

(2) This paper reports on an application that delivers automated formative feedback designed to help university students improve their assignments. (3) The aim of the system is to improve the confidence and skills of the user by promoting self-directed learning through metacognition. (4) The system focuses on the content of an essay by using automatic summarisation techniques, automatic structure recognition, diagrams, animations, and interactive exercises that promote reflection. (15) The system is currently undergoing initial exploratory rounds of testing by ex-student volunteers and will be the subject of two full-scale empirical evaluations starting in September 2013. (1) The main claims of this paper are the application and adaptation of graph-based key word and key sentence ranking methods for a novel purpose, and ensuing observations concerning the suitability of two different centrality algorithms for the purposes of key word extraction.

1 Introduction

A fundamental problem in distance education is student attrition, particularly during the early months, largely due to low morale. Graduation rates at distance-learning institutions are often less than evident at the level of individual modules or course units, where completion rates may be as low groups of students, such as those from ethnic minorities (Richardson, 2012). Some students who have dropped out of open university courses have reported that the reason they left was a conviction of their own in adequacy when assignments. These reports are backed up by the drop-out rate that occurs just before the first courses, is typically as high as 30%.
Sample key phrases dispersion plot
Short text for illustration of Rainbow Diagrams

Text (Extract from online FAQ about foxes)
Will the foxes in my garden attack my dog or cat?
This is extremely unlikely. 2. **Foxes avoid dogs, even small dogs, because many foxes are killed by dogs.**
So it is much more likely that your dog will attack the fox, not the other way round. Attacks on cats are equally rare: cats and foxes are roughly the same size, and cats are very capable of defending themselves against foxes. So it is hardly surprising that foxes generally give cats a wide berth and flee when threatened by a cat. Occasionally small kittens are killed, but this is rare. Keeping your cat indoors at night greatly reduces the chances of an encounter with a fox. 8. **There are also a variety of other benefits: cats kept in at night are healthier and live longer, and kill less of the local wildlife.**
Sentence graph of short text

Last sentence (node 8 - red)
"There are also a variety of other benefits: cats kept in at night are healthier and live longer, and kill less of the local wildlife"

Second sentence (node 2 - violet)
"Foxes avoid dogs, even small dogs, because many foxes are killed by dogs."
Pretend essay: 10 identical paragraphs
Pretend essay: 50 identical sentences
Stanford University Boothe Prize essay

DMW AHE Conference June 2017
OU essay awarded high grade
OU essay awarded low grade
Rainbow diagrams related to mark awarded

- Multivariate analysis of variance on marks awarded to 45 students
- Submitted two essays
- Rainbow diagrams produced from these essays and rated as high, medium or low attainment
- Covariate showed a significant relationship with the marks
- $F(1, 43) = 5.92, p = .01$ using a directional test
- Essays rated as high would be expected to receive 8.56 percentage points more than essays rated as medium
- 17.2 percentage points higher than essays rated from rainbow diagrams as low
How about feedback first?

- Hints before writing?
- R.C.T.
- 2 essays
- $F(1,41) = 3.23 \ p = 0.04$ for hints

DMW AHE Conference June 2017
How about anywhere anytime testing?
TeSLA: Adaptive Trust based e-Assessment

- AIM: Secure & reliable online assessment
- TECHNOLOGIES: Voice/face recognition, keystoke pattern detection, anti-plagiarism and forensic analysis
- 18 European partners, OUUK responsible for evaluation

http://tesla-project.eu/
Creating teaching and learning dialogues: towards guided learning supported by technology

- Learning to judge
- Providing reassurance
- Providing a variety of signposted routes to achieve learning goals
- Provide socio-emotive support
References


References (2)


References (3)


References (4)
